Universidade de São Paulo Instituto de Biociências

A duração das linhagens e a dinâmica macroevolutiva de Ruminantia (Mammalia)

Lineage duration and macroevolutionary dynamics of Ruminantia (Mammalia)

Matheus Januario Lopes de Sousa

São Paulo 2019 Universidade de São Paulo Instituto de Biociências Programa de pós-Graduação em Ecologia

Matheus Januario Lopes de Sousa

A duração das linhagens e a dinâmica macroevolutiva de Ruminantia (Mammalia)

Lineage duration and the macroevolutionary dynamics of Ruminantia (Mammalia)

Dissertação apresentada ao Instituto de Biociências da Universidade de São Paulo como parte dos requisitos para obtenção do título de Mestre em Ciências, na área de Ecologia de Ecossistemas Terrestres e aquáticos.

Orientador: Tiago Bosisio Quental

Versão corrigida

Cópia original encontra-se disponível no instituto de Biociências da USP

São Paulo

Maio, 2019

Ficha catalográfica

Sousa, Matheus Januario Lopes de

A duração das linhagens e a dinâmica macroevolutiva de Ruminantia (Mammalia) . Versão em inglês: Lineage duration and macroevolutionary dynamics of Ruminantia (Mammalia)/ Matheus Januario Lopes de Sousa; orientador Tiago Bosisio Quental. -- São Paulo, 2019.

106 p.

Dissertação (Mestrado) - Instituto de Biociências da Universidade de São Paulo, Departamento de Ecologia.

Palavras-chave: 1. Registro fóssil. 2. Extinção. 3. Hipótese da rainha vermelha. 4. Paleontologia. 5. Ecologia evolutiva.

Universidade de São Paulo. Instituto de Biociências. Departamento de Ecologia

Comissão julgadora:

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof. Dr. Tiago bosisio Quental

"Progress is made by young scientists who carry out experiments that old scientists said wouldn't work" Frank Westheimer

> "O trabalho [científico] nunca termina, o que termina é o prazo" Maria Luísa Jabbur

About the covers

The connections between doing science and doing art are many. But a clear divisor between them is that science does not aim to put emotions in its final product. I really tried to get my emotions out of this dissertation, but I still felt that I, in some way, should put my emotional connection to the questions investigated here. So, my choice was doing so through the covers. All the covers are my re-readings of the same work of art made by Eshel Ben-Jacob (1952 – 2015), a physicist which used microbial communication under stressful Petri dishes to generate beautiful complex patterns in bacterial culture. Even with his passing, his works are still available in www.microbialart.com. Each of my re-readings represents my emotional perception of what was done in each part of this dissertation.

The main cover is a mixture of noise and pattern and is related to how I perceive macroevolution as a whole. It undeniably shows us intense trends at its core and points into the direction of something. But many of the connections, especially the ones which are far from the core, are still unknown, and looking at them compels me to search for these patterns. To do so I can zoom in. I can zoom out. I can compare contrasts, and I can do what I've done in each chapter: the emphasis in the points, and not in the temporal patterns, can neglect a bit the processes that happened, but it is still an important step if one wants to be sure that a point in the cover is, in fact, a point. Before a step further, sometimes we need to step back. This is what motivated chapter one, and so it's cover. The cover of the second chapter represents the application of a model (the black circle) in the hope it can help us to see patterns more clearly but also left us with questions about possible artifacts that came from the application of the model. This is what was done in chapter 2, and was visually represented in its cover. All the covers also relate visually the birth-death process of these bacteria observed through modern scientific techniques and equipment to the birth-death process observed in the macroevolutionary scale, with a spectacular and unintentional visual reference to the classic diagrams made by paleontologists in the begging of the past century. This is so awesome. Finally, I must admit that I like the irony of the macroevolution of life being represented with microorganisms.

,			
T	1	•	
- n		16	0
	u	•	
			-

Resumo	8
Abstract	9
Introdução geral	10
Referências	14
Capítulo 1: Ruminantia macroevolutionary patterns are robust	to distinct taxonomic
treatments on data from paleontological online databases	17
Abstract	18
Introduction	19
Material and Methods:	21
Data collection and curatorial work	21
Birth-death diversification analysis	22
Comparing different diversification inferences	24
Results	24
Direct comparisons between the datasets	24
Birth-death model analysis	25
Discussion	32
References	37
Supplementary material	40
Supplementary figures (printed version)	40
Supplementary file 1	50
Supplementary file 2	53
Supplementary file 3	56
Supplementary figures (digital version)	57
Supplementary tables	60
Capítulo 2: Re-evaluation of the "law of constant extinction" for	ruminants at
different taxonomical scales	61
Abstract	62
Introduction	63
Methods:	65
Results:	68
Discussion:	74
References:	79
Supplementary material	84
Supplementary file 1	84
Supplementary figures	86
Conclusão geral	96
Referências	98

Agradecimentos

Se conhecimento é poder, o pouco que (gosto de pensar que) tenho veio e todos os dias continua vindo graças à ajuda e paciência de muita gente. Mais gente do que caberia em todas as páginas dessa dissertação. Os dois maiores culpados certamente são meus pais. Meu pai sempre fez questão de fomentar minha vontade de aprender: lembro dele lendo curiosidades sobre dinossauros para mim desde quando eu era menos inteligente que um corvo, e também lembro dele passar minha infância e adolescência me mostrando de forma prática que após a comida, a necessidade humana mais importante é o conhecimento. Em adição ao platonismo do meu pai, minha socrática mãe sempre me deu confiança de que era possível buscar minha eudaimonia, me dando lições contínuas de como se virar no mundo, mas sem dúvida a principal contribuição dessa mulher incrível é nunca ter permitido que eu desistisse de mim mesmo. Mãe e pai, vocês são realmente uma dupla incrível e eu não seria um pedaço do que sou sem o cuidado parental de vocês.

Essa dissertação jamais teria sido escrita sem a ajuda direta de muita gente. A primeira delas é o Tiago. Sou muito grato por você ter acolhido e depois me ajudado a esculpir minhas ideias mais malucas, mas principalmente por você respeitar minha personalidade debatedora, meu tempo e minhas escolhas de aprendizado e de trabalho. Sério Chefe, espero um dia herdar sua exemplar forma de orientar. Outra pessoa com uma parcela gigantesca de culpa por minha dissertação estar finalizada é o Ari, cuja paciência pra ensinar é (de forma não intencional, eu juro) testada por mim a pelo menos 5 anos. Sem suas explicações e muita ajuda com os problemas cabeludos eu estaria condenado a sempre usar *loops*, como se eu fosse um homem das cavernas, e a finalização da minha dissertação certamente demoraria até a próxima era do gelo. Lóra e Graxa, eu me sinto muito sortudo por ter coexistido com companhias tão leves e divertidas! Suas ideias estarão em minha cabeça, mas vocês sempre estarão em meu coração. Se aprendi algo no LabMeMe, é que não se deve esquecer das linhagens-irmãs e nem das linhagens que se "extinguiram" pelo caminho: Sem-Vê, Toshiba, Mathias, Gentil, Isabel e Frodo, a minha convivência menor com vocês não lhes impediu de me ensinar e me influenciar muito, no que tange assuntos acadêmicos ou não. LabMeMers, novamente tenho que agradecer a todos vocês que trabalharam comigo durante essa dissertação e antes dela. Sou muito feliz por ter germinado

minha carreira neste laboratório, mas sou muito mais grato pelo fato do laboratório ter me permitido conviver com pessoas tão gentis e maravilhosas.

Gosto de pensar que os estudantes são o resultado de quem os ensina. E eu não poderia deixar de agradecer aos meus grandes tutores: Dani, "obrigado" é uma palavra insuficiente para te agradecer por toda a metacognição que aprendi com você e especialmente por você me dar a espinha dorsal da minha formação como professor. Miúdo, obrigado pelos comitês elucidantes após os cafés, pelas aulas deslumbrantes após os almoços e pelas cervejas geladas após o fim das semanas. Edu, muito obrigado pelas excelentes sugestões no comitê e pela disciplina de meta-análise, que mudaram radicalmente a organização desta dissertação. Alê e P.I., vocês basicamente me deram as ferramentas (e as oportunidades de aprender a usá-las e avaliá-las) para entender e esmiuçar as evidências (ah, as evidências...), então sou muito grato por ter sido estudante-monitorado e também estudante-monitor durante suas disciplinas (e na de muitos outros professores, muito obrigado pela oportunidade!). Sendo sincero, eu perdi a contagem das noites que passei em claro maravilhado ou pensativo sobre o que aprendi nos cursos do Márcio Martins, Adriana Martini, Camila Castanho, Ricardo Sawaya, Roberto Kraenkel e Renato Coutinho, além dos grandes professores já citados anteriormente. Muito obrigado a todos.

Agradeço também a todos as trabalhadoras e trabalhadores do CNPq, da USP e do IB (em especial para o pessoal dos bandejões, bibliotecas, manutenção, portarias, secretarias e corpo técnico), por proporcionarem a estrutura e o ambiente que me permitiu estudar, trabalhar e me deu uma comunidade. Um agradecimento especial à Vera, à Erika e à Shirlene, por terem me ajudado de forma tão solícita e eficiente durante todos os apuros burocráticos em que me meti.

Ao Ushuaia, meu grande parceiro de descobertas, eu agradeço por tudo. Ahazo, minha grande parceira de discussões, você me ajuda tanto que "obrigado" já é repetitivo. Blue, obrigado pelos scripts em linguagens que eu não entendo. À Mar, eu agradeço por ter sido companhia no amor e na dor que foram 2017 e 2018. Me falta espaço, mas felizmente me sobram amigos. Então quem não foi citado ainda eu agradecerei em nossos próximos encontros, calorosa e pessoalmente.

Muito obrigado a todos pela ajuda na jornada. Do fundo do meu coração.

Resumo

O objetivo geral desta dissertação é testar hipóteses relacionadas à duração das linhagens (i.e. o intervalo de tempo definido pela origem e extinção das diferentes linhagens) e à dinâmica da diversificação dos ruminantes. Para isso, nós usamos compilações públicas de dados fósseis e modelos probabilísticos. Como a literatura macroevolutiva se aproveita das múltiplas escalas taxonômicas onde é possível estudar a mudança da biodiversidade no tempo, os dois capítulos desta dissertação testam suas hipóteses no níveis de espécie e gênero na tentativa de iluminar o papel da escala na manifestação dos padrões macroevolutivos.

No primeiro capítulo, nós investigamos se tratamentos taxonômicos distintos aplicados aos dados brutos alteram os resultados obtidos por um método macroevolutivo. Nossos resultados sugerem que os dados, mesmo após passar por diferentes tratamentos, indicam similares dinâmicas de diversificação tanto no nível taxonômico dos gêneros quanto no nível das espécies. Por outro lado, as comparações entre dois conjuntos de dados inicialmente distintos podem gerar diferenças consideráveis na dinâmica da diversificação e inclusive indicar diferenças entre escalas taxonômicas. Esses resultados são robustos mesmo que as imperfeições do registro fóssil sejam consideradas de diferentes maneiras pelo método usado.

No segundo capítulo, nós usamos modelos probabilísticos para testar uma importante hipótese macroevolutiva que é também considerada uma premissa para vários estudos da área: a de que a probabilidade de extinção de qualquer linhagem é independente da sua longevidade, ou seja, do intervalo de tempo que a linhagem já viveu. Nossos resultados sugerem um padrão consistente de diminuição na probabilidade de extinção conforme as espécies vivem por mais tempo, enquanto que o padrão observado do no nível dos gêneros é menos claro. Este resultado não só enriquece o nosso conhecimento a respeito da dinâmica da extinção de linhagens, como sugere que a premissa de independência de idade na probabilidade de extinção (importante para vários estudos macroevolutivos) pode não ser valida, pelo menos para Ruminantia.

Nós esperamos que os dados e ideias aqui expostos interajam com novas ideias e dados do leitor, estimulando o desenvolvimento da macroevolução.

Abstract

The main goal of this dissertation is to test hypotheses related to the duration of the lineages (i.e. the time interval between the origin and extinction of each lineage) and to the diversification dynamics of the ruminants. To achieve this goal, we used public compilations of fossil data and probabilistic models. As the macroevolutionary literature take advantage of multiple taxonomic scales in which is possible to study biodiversity changes through time, the two chapters of this dissertation tested different hypotheses in both the species and genus levels, as an attempt to shed some light on the role of scale influencing the macroevolutionary phenomena.

In the first chapter, we investigated if different taxonomic treatments applied over the raw data significantly change the results obtained from a macroevolutionary method. Our results suggest that the same data, even passing through different data curations, indicate similar diversification dynamics in species and genus taxonomic levels. On the other hand, data that comes from different sources may present considerable differences in the diversification dynamics, and even indicate differences between taxonomic scales. These results are robust to different considerations of the preservation biases of the fossil record by the method.

In the second chapter, we used probabilistic models to test an important macroevolutionary hypothesis that is also considered a premise of several macroevolutionary studies: that the probability of extinction of any lineage is independent of its longevity (the time span determined by the origination and extinction of each lineage). Our results suggest a consistent pattern of decreasing extinction probability as species live for longs periods of time, but the pattern presented by the genera is less clear. These results not only enhance our knowledge about extinction dynamics, but it also suggests that the premise of age extinction probability for independence in the (important several macroevolutionary studies) might not be valid, at least for Ruminatia.

We hope our data and ideas interact with new ideas and data from the reader to stimulate the development of the macroevolutionary field.

Introdução geral^{*}

"Uma das primeiras "regras da biodiversidade" a serem investigadas diz respeito à idade e ao risco de extinção: seriam as linhagens "anciãs" mais ou menos propensas a sobreviver que as linhagens que recentemente emergiram? Se isso puder ser resolvido, então haverá meios de estimar fatores de risco de extinção das espécies viventes"

Michael Benton (2016) [tradução livre]

Ao observarmos a enorme variação presente na biodiversidade, o primeiro padrão que surge é extremamente curioso: esta variação não está distribuída de maneira uniforme. Por mais que a biodiversidade exiba uma miríade de formas, funções e interações, cada uma dessas propriedades está distribuída de forma desigual no espaço, no tempo e na árvore da vida. Existem formas e funções mais frequentes na natureza, além de interações químicas, fisiológicas, genéticas e ecológicas mais comuns, e existem, claro, as exceções às "regras". Tentar entender ao menos uma parte dos processos que geraram e selecionaram as variantes ao longo de milhões de anos, e descrever a história da vida, que resultou na variação que podemos observar na natureza, contemporânea ou passada, são as duas primeiras preocupações da macroevolução, área de meu interesse pessoal e onde esta dissertação está situada. A primeira de todas as manifestações dessa variação natural que a macroevolução estuda é o número (riqueza) de espécies. Esta variação da riqueza é o resultado de dois processos, que agem simultâneos no tempo: a especiação (que adiciona espécies ao longo do tempo) e a extinção (que remove espécies ao longo do tempo. O resultado da interação entre especiação e extinção é o que chamamos de diversificação líquida (ou apenas diversificação). As taxas com que esses dois eventos ocorrem são influenciadas por eventos históricos, como extinções em massa (Jablonski 1986; 2017) ou intercâmbios faunísticos (como o intercâmbio americano - Stehli & Webb, 2013), por características dos organismos, como amplitude da distribuição geográfica (Jablonski 2008; 2017) ou inovações-chave (Hunter, 1998), por fatores bióticos (Ezard et al 2016; Silvestro et al 2015), e por mudanças no ambiente (Benton, 2009; Mayhew et al, 2012).

Para realizar investigações sobre esses temas, a macroevolução usa duas vias principais de evidência: as filogenias moleculares e o registro fóssil. De um lado, as filogenias moleculares são amplamente utilizadas em grupos viventes, principalmente quando o registro fóssil desses grupos não é muito bem amostrado e as relações filogenéticas entre as espécies são relativamente bem conhecidas. Entretanto, filogenias moleculares são limitadas no sentido de que não fornecem informação direta sobre a extinção (Quental & Marshall, 2010; Rabosky 2010). Por outro lado, o registro fóssil fornece informação direta sobre a especiação e a extinção (Quental & Marshall, 2010). Apesar desta vantagem, o registro fóssil é imperfeito, possuindo diferenças na

^{*}As referências da introdução e conclusão gerais estão formatadas de acordo com as normas da revista Palebiology

qualidade da preservação tanto no espaço quanto no tempo, que dificultam ou até mesmo impedem que os fósseis sejam interpretados como evidência direta de ocorrência temporal de organismos para uma série de grupos taxonômicos. Cabe ressaltar que a datação de um organismo é tipicamente feita a partir do intervalo de tempo da sessão estratigráfica na qual se encontra o fóssil, e não a partir de uma datação direta do fóssil. Esses diferentes aspectos fazem com que o uso de dados fósseis normalmente requeira uma quantidade grande de dados (ocorrências temporais de organismos), o que torna o registro fóssil útil para essa finalidade apenas para alguns grupos de organismos com um registro fóssil razoavelmente rico, como por exemplo moluscos e mamíferos. Quando possível, a integração das duas vias de evidência (filogenias moleculares e registro fóssil) aumenta o poder de resposta, mas o uso integrado das duas, apesar de ser discutido na literatura a muito tempo (Quental & Marshall, 2010; Morlon *et al*, 2011) ainda é algo relativamente inexplorado de forma explícita e diretamente integrada na macroevolução.

Se a macroevolução é, resumidamente, estudar padrões de mudança na biodiversidade em longas escalas de tempo, e a mais simples medida de biodiversidade é a riqueza (ou seja: número) de espécies, é natural que seja relevante atentarmos para como as espécies são definidas e descritas. A discussão sobre o que é uma espécie "válida" para se estudar evolução é extensa na literatura biológica (como demonstrado no primeiro capítulo do clássico livro de Coyne & Orr (2004)), e fundamental na geração de teoria macroevolutiva (Coyne & Orr, 2004; Rabosky, 2016), mas decidir que tipos de espécies (e portanto de dados) são desejados numa análise macroevolutiva vai além da discussão teórica, e encontra aspectos práticos que normalmente não são abordados diretamente na literatura. Dois aspectos importantes são como a quantidade e a qualidade dos dados podem influenciar as estimativas dos padrões da biodiversidade ligados a amplas escalas temporais, espaciais e taxonômicas.

A compilação de dados fósseis com o intuito de estudar padrões de diversidade é um exercício antigo por parte dos paleontólogos (e.g. Newell 1963), apesar da abordagem geralmente sofrer uma série de críticas. Além das críticas usuais referentes à qualidade do registro fóssil (Liow et al, 2011; Silvestro et al, 2014a), essas críticas também se referem à construção, manutenção, e uso indiscriminado das próprias bases de dados (e.g. Stearn, 1999; Prothero, 2014). Parte dessas críticas são bem fundamentadas: guando uma grande guantidade de dados é compilada e organizada, é comum que uma série de decisões ad hoc sejam tomadas e nem sempre explicitadas ou sequer documentadas. Isso, aliado à inerente possibilidade de falha humana, torna grande parte das compilações irreproduzível em essência e inverificável dada a sua magnitude. Um outro importante fator é que o conhecimento científico avança e as bases de dados podem não acompanhar essas mudanças (e.g. reclassificações taxonômicas) em tempo satisfatório, de forma que algumas informações podem permanecer desatualizadas por tempo considerável. Dadas todas essas questões, é surpreendente que as conclusões geradas a partir de dados como o compêndio de Joseph John "Jack" Sepkoski (1982), tenham se mostrado robustas às atualizações científicas, e tenham se mantido essencialmente intactas mesmo que o compêndio tenha passado por recompilações (Sepkoski, 1992) e reavaliações internas (Sepkoski, 1993) e

externas (Adrain & Westrop, 2000) (para outro exemplo dessa estabilidade, veja Maxwell & Benton, 1990).

Hoje não é mais necessário que uma única pessoa compile a quantidade exorbitante de dados que Sepkoski ou outros paleontólogos fizeram. Vivemos num mundo onde uma grande quantidade de dados biológicos está disponível a virtualmente todos os computadores do mundo, sendo que os dados fósseis sobre mamíferos estão alocados principalmente em dois deles: o Paleobiology Database (PBDB - <u>http://paleobiodb.org/</u>), e o New and Old Worlds Database (NOW - <u>http://www.helsinki.fi/science/now/</u>). Apesar de seu altíssimo valor, essas bases de dados modernas não estão imunes a uma série de problemas (geralmente por conta de falha humana durante a entrada de dados - Zizka *et al*, 2019). Desta forma, mesmo as mais modernas bases de dados não devem ser utilizadas diretamente sem que haja algum tipo de revisão dos dados antes de uma análise. No primeiro capítulo serão abordados diferentes níveis de curadoria de dados, incluindo algumas muito simples que não requerem conhecimento taxonômico do grupo de interesse, podem alterar as conclusões de análises macroevolutivas interessadas em estimar taxas de diversificação. Para tal, serão utilizados dados da linhagem Ruminantia (Mammalia) obtidos nas bases de dados PBDB e o NOW.

No segundo capítulo, a lei da extinção constante, proposta por Van Valen (1973), será reavaliada utilizando os dados curados no primeiro capítulo. Em 1973, Van Valen apresentou resultados que indicavam que toda a biodiversidade global seguia um mesmo padrão de extinção: táxons (Van Valen usou principalmente gêneros, mas também famílias e em raros casos, espécies) dentro de um mesmo grande grupo (e.g. espécies dentro de gêneros ou de famílias) possuíam uma mesma probabilidade de extinção ao longo de diferentes idades, o que significaria que o tempo que um táxon já viveu não informa sobre a probabilidade do mesmo se extinguir. A ampla consistência do resultado de Van Valen em praticamente todos os grandes grupos biológicos com dados disponíveis na época não apenas influenciou a criação da hipótese da rainha vermelha para explicar esse padrão de taxas constantes (descrita no mesmo artigo de Van Valen, em 1973), como também justificou empiricamente a premissa de independência de idade na probabilidade de extinção que foi posteriormente incorporada em basicamente todos os modelos macroevolutivos de nascimento e morte (Nee et al, 1994). Apesar da grande relevância teórica dos resultados de Van Valen (1973), reavaliações posteriores no nível de espécie sugerem o oposto, e somente o trabalho de Pearson (1995) com nanoplâncton calcário encontrou evidência do padrão descrito por Van Valen, em que a probabilidade de extinção é independente da idade de um táxon. A grande parte das outras reavaliações encontrou que a probabilidade de extinção aparenta decrescer (Boyajian, 1986; 1991; Crampton et al, 2016, Hagen et al, 2018) ou aumentar (Pearson, 1995; Doran et al, 2006; Finnegan et al, 2008; Ezard et al, 2016) com a idade da linhagem. O número limitado de posteriores reavaliações do padrão de extinção encontrado por Van Valen (1973) pode ser explicado principalmente pela imperfeição da preservação e consequentes vieses inerentes ao registro fóssil (Liow et al, 2011; Hagen et al, 2018) que permitem que apenas um número pequenos de clados possa ser reavaliado no nível de espécie com alguma confiabilidade. Recentemente, um novo método (Hagen et al, 2018), que

lida direta e explicitamente com essas imperfeições, permitiu a possibilidade de reavaliação dos padrões propostos por Van Valen (1973) para uma ampla variedade de clados e em diferentes níveis taxonômicos, incluindo os relativos às espécies e aos gêneros dos ruminantes, grupo focal de estudo desta dissertação.

Em ambos capítulos serão analisadas a dinâmica macroevolutiva em dois níveis taxonômicos diferentes: gênero e espécie. Dada a incompletude do registro fóssil, não é incomum que sejam encontrados poucos fósseis de uma mesma espécie, ou que os espécimes encontrados não possam ser identificadas no nível de espécie. Assim, a literatura paleontológica comumente utiliza dados a nível de gênero, assumindo que o padrão de gêneros é, ao menos em algum nível, análogo ao de espécies, ao mesmo tempo em que é relativamente melhor amostrado. O uso de gêneros (ou outros níveis taxonômicos superiores) como unidades de interesse macroevolutivo foi adotado amplamente, por exemplo, no próprio trabalho de Van Valen (1973). Nesta dissertação irei portanto investigar outra prática comum em estudos paleontológicos, o uso de níveis taxonômicos mais altos como uma boa aproximação para a dinâmica das espécies. Apesar da prática paleontológica, deve-se perguntar se de fato diferentes níveis taxonômicos apresentam dinâmicas análogas, e se dada a quantidade de dados e o poder dos métodos disponíveis nos dias atuais ainda é necessária a utilização de dados a nível de gênero se o objetivo é, no fim, chegar a padrões relativos à dinâmica evolutiva das espécies. Isso é especialmente relevante já que modelos matemáticos desenvolvidos para se estudar a relação entre longevidade e probabilidade de extinção (Raup, 1978) e poucas descrições empíricas (por exemplo Ezard et al, 2016) sugerem padrões distintos entre os dois níveis taxonômicos.

Por fim, a escolha dos ruminantes (Ruminantia, Mammalia) como objeto de estudo representa tanto uma escolha pragmática quanto um caminho natural dados os meus interesses na graduação por alguns grupos de ruminantes. Desta forma, as perguntas respondidas nos dois capítulos são mais referentes à teoria macroevolutiva do que ao grupo em específico. Assim, a dissertação se aproveita de características dos ruminantes que levaram às questões aqui abordadas. A primeira dessas características é que o grupo possui uma taxonomia relativamente estável para as espécies viventes (Cantalapiedra et al, 2014), algo importante para o método que empregamos nos dois capítulos (Silvestro et al, 2014a; Silvestro et al, 2014b; Silvestro et al, 2018a). A segunda delas é que a taxonomia fóssil parece menos estável que a das espécies viventes, porém passando por mudanças e importantes revisões nas últimas décadas (Prothero, 2014), o que permite que nem todas as mudanças discutidas na literatura tenham sido já incorporadas nos bancos de dados que utilizamos. Essas características permitem que haja espaço para um revisão taxonômica dos dados, como a que fizemos para o primeiro capítulo desta dissertação. A revisão taxonômica feita no capítulo 1, por sua vez, pode melhorar a qualidade dos dados usados no capítulo 2, além de permitir que seja reduzida a possibilidade de pseudo-especiação (onde uma mesma linhagem muda anageneticamente no tempo o suficiente para ser reconhecida pela literatura fóssil como duas ou mais espécies) e pseudo-extinção afetarem o nosso conjunto de dados, e portanto nossas estimativas das "verdadeiras" idades das linhagens. Essas revisões se mostram relevantes, pois a duração das linhagens possui uma

relação matemática direta com o comportamento temporal da extinção do clado (Marshall, 2017), e a literatura fóssil a um tempo prevê (Prothero, 2007; Prothero, 2014) e recentemente encontrou evidência (Silvestro 2018b) de que a pseudo-especiação e pseudo-extinção podem ser um fator relevante na diversificação dos ruminantes. Apesar da generalização desse efeito ser difícil para o resto da biodiversidade (veja Strotz *et al*, 2013) e que a real influência da pseudo-especiação e pseudo-extinção só possam ser efetivamente estimadas com uma abordagem filogenética explícita, a revisão taxonômica aqui realizada já lida com alguns desses problemas relacionados à anagênese. Além dos motivos apontados, os ruminantes possuem um registro fóssil rico em famílias, gêneros e espécies (Prothero, 2008; Cantalapiedra 2015) e uma distribuição geográfica concentrada na região focal dos dois bancos de dados utilizados na dissertação (PDBD e NOW). Assim, os ruminantes provêm dados adequados para o estudo da duração de linhagens, incluindo o efeito da taxonomia e do nível de organização biológica (nível taxonômico), que são objetivos centrais desta dissertação.

Referências

- Adrain, J. M. and S. R. Westrop 2000: An Empirical Assessment of Taxic Paleobiology. Science 289:110-112.
- Bailey, N. T. J. 1964. The Elements of Stochastic Processes. 249 pp. John Wiley & Sons, Inc..: New York, NY.
- Benton, M. J. 2009. The Red Queen and the Court Jester: species diversity and the role of biotic and abiotic factors through time. science 323: 728-732.

Benton, M. J. 2016: Origins of biodiversity. Palaeontology 56:1-7.

- Boyajian, G. F. 1991: Taxon Age and Selectivity of Extinction. Paleontological Society 17:49-57.
- . 1986: Phanerozoic trends in background extinction: consequence of an aging fauna. Geology 14:955–958.
- Cantalapiedra, J. L., M. Hernández Fernández, B. Azanza, and J. Morales. 2015: Congruent phylogenetic and fossil signatures of mammalian diversification dynamics driven by Tertiary abiotic change. Evolution 69:2941–2953.
- Cantalapiedra, J. L., R. G. FitzJohn, T. S. Kuhn, M. H. Fernández, D. DeMiguel, B. Azanza, J. Morales, and A. Mooers. 2014: Dietary innovations spurred the diversification of ruminants during the Caenozoic. Proceedings of the Royal Society B: Biological Sciences 281.
- Crampton, J. S., R. A. Cooper, P. M. Sadler, and M. Foote. 2016: Greenhouse-icehouse transition in the Late Ordovician marks a step change in extinction regime in the marine plankton. Proceedings of the National Academy of Sciences 113:1498–1503.

Coyne, J. A., and H. A. Orr, 2004. Speciation. Sinauer. Sunderland, MA.

Doran, N. A., A. J. Arnold, W. C. Parker, and F. W. Huffer. 2006: Is extinction age dependent? Palaios 21:571–579.

Ezard, T. H. G., T. B. Quental, and M. J. Benton. 2016: The challenges to inferring the regulators of biodiversity in deep time. Philosophical Transactions of the Royal Society B: Biological Sciences 371:20150216.

Finnegan, S., J. L. Payne, and S. C. Wang. 2008: The Red Queen revisited: reevaluating the age selectivity of Phanerozoic marine genus extinctions. Paleobiology 34:318–341.

- Hagen, O., T. Andermann, T. B. Quental, A. Antonelli, and D. Silvestro. 2018: Estimating Age-Dependent Extinction: Contrasting Evidence from Fossils and Phylogenies. Systematic Biology 67:458–474.
- Liow, L. H., L. Van Valen, and N. C. Stenseth. 2011: Red Queen: From populations to taxa and communities. Trends in Ecology and Evolution 26:349–358.
- Marshall, C. R. 2017: Five palaeobiological laws needed to understand the
- Mayhew, P. J., M. A. Bell, T. G. Benton, and A. J. McGowan. 2012. Biodiversity tracks temperature over time. Proceedings of the National Academy of Sciences 109: 15141-15145.
- Maxwell, W. D., M. J. Benton, and N. Summer. 2007: Historical Tests of the Absolute Completeness of the Fossil Record of Tetrapods Historical tests of the absolute completeness of the fossil record of tetrapods. 16:322–335.
- Nee, S., R. M. May, and P. H. Harvey. 1994: The reconstructed evolutionary process. Philosophical Transactions of the Royal Society B: Biological Sciences 344:305–311.
- Newell, N.D. 1963. Crises in the history of life. Scientific American 208:76-92
- Pearson, P. N. 1995: Investigating age-dependency of species extinction rates using dynamic survivorship analysis. Historical Biology 10:119–136.

Prothero, D. R., & Foss, S. E. (Eds.). 2007. The evolution of artiodactyls. JHU Press. Baltimore, MD

- Prothero, D. R. 2008: Systematics of the musk deer (Artiodactyla: Moschidae: Blastomerycinae) from the Miocene of North America. New Mexico Museum of Natural History and Science Bulletin 44:207– 223.
- -----. 2014: Garbage in, garbage out: The effects of immature taxonomy on database compilations of

North American fossil mammals. New Mexico Museum of Natural History and Science Bulletin 68:257–264.

- Rabosky, D. L. 2016: Reproductive isolation and the causes of speciation rate variation in nature. Biological Journal of the Linnean Society 118:13–25
- Raup, D. M., and N. Winter. 1978: Cohort Analysis of Generic Survivorship Cohort analysis of generic survivorship. Analysis 4:1–15.
- Sepkoski, J. J. 1982: A compendium of fossil marine animal families. Contributions in biology and geology 52.
- Sepkoski, J. J. 1992: A compendium of fossil marine animal families, 2nd edition. Contributions in biology and geology 83:1–156.
- Sepkoski, J. J. 1993: Ten years in the library : new data confirm paleontological patterns.Paleontological Society. 19:43–51.
- Silvestro, D., N. Salamin, and J. Schnitzler. 2014a: PyRate: A new program to estimate speciation and extinction rates from incomplete fossil data. Methods in Ecology and Evolution 5:1126–1131.
- Silvestro, D., J. Schnitzler, L. H. Liow, A. Antonelli, and N. Salamin. 2014b: Bayesian estimation of speciation and extinction from incomplete fossil occurrence data. Systematic Biology 63:349–367.
- Silvestro, D., A. Antonelli, N. Salamin, T. Quental. 2015: The role of clade competition in the diversification of North American canids. Proceedings of the National Academy of Sciences 112:8684-8689.
- Silvestro, D., A. Antonelli, N. Salamin, and X. Meyer. 2018a: Improved estimation of macroevolutionary rates from fossil data using a Bayesian framework. BioRxiv 18:316992.
- Silvestro, D., R. C. M. Warnock, A. Gavryushkina, and T. Stadler. 2018b: Closing the gap between palaeontological and neontological speciation and extinction rate estimates. Nature Communications 9.
- Strotz, L. C., and A. P. Allen. 2013. Assessing the role of cladogenesis in macroevolution by integrating fossil and molecular evidence. Proceedings of the National Academy of Sciences 110: 2904-2909.

Van Valen, L. 1973: A New Evolutionary Law. Evolutionary Theory 1:1-30.

Zizka, A., D. Silvestro, T. Andermann, J. Azevedo, C. Duarte Ritter, D. Edler, H. Farooq, A. Herdean, M. Ariza, R. Scharn, S. Svantesson, N. Wengström, V. Zizka, and A. Antonelli. 2019: CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. Methods in Ecology and Evolution

Capítulo 1

Ruminantia macroevolutionary patterns are robust to distinct taxonomic treatments on data from paleontological online databases*

Matheus L. Januario¹ & Tiago B. Quental^{1,}



1. Departamento de Ecologia, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil

[·] Manuscrito formatado de acordo com as normas da revista Paleobiology

Abstract

One of Macroevolution goals is to describe changes in biodiversity in deep time. The fossil record can be used for this endeavor but performing this type of analyses requires large amounts of data. The creation of online databases has clearly promoted this approach, but errors are inherent to any kind of database, and fossil datasets are no exception. Those errors range from taxonomic inconsistencies to inputting mistakes that might affect analysis further down the line. The use of genera instead of species as the biological unity of analysis is one of the practices implemented to include a large number fossil data in macroevolutionary analyses, but the direct correspondence between genera and species level is debated in literature. Here, we examine, at the species and genus level, the impact of different levels of data curation, including different taxonomic treatments, in our inferences in diversification dynamics of ruminants using a Bayesian framework. We found that curatorial work in the same dataset, as expected, impacted diversification dynamics estimates, but that simple procedures such as code-oriented curatorial work already resulted in similar inferences to those of more detailed and time-consuming taxonomic reviews. The reason for such automatic curatorial efficiency is because fossil databases, in particular the Paleobiology Database, have enough information onthemselfs to inform potential inadequacies of the data. On the other hand, we showed that the two most curated datasets, one using our original dataset and another done by other investigators using a partially different dataset, showed considerable differences in the species level analysis. Exaplaining this difference is not straitgh forward because unique species in the dataset build by other researches are concentrated in the time window where the diversification dynamics were quite congruent between our most curated dataset and the other researcher's dataset. We also note that the most strinking differences are concentrated where the other dataset has a considerably lower occurrence temporal resolution. We also show that different taxonomic practices and data choices might strongly influence how good of a proxy are genus level analysis are if one is trying to infer species level dynamics: in our most curated dataset genus level is a good proxy for species, but there is smaller correspondence between the two taxonomic levels in other published dataset of the same group. We also hope the framework presented here contributes to a more data-driven and reproducible curatorial practice used in macroevolutionary studies based on publicly available paleontological data.

Key-words: Data quality, Fossil record, Paleobiology Database, New and Old Words Database, Taxonomic surrogacy

Introduction

The asymmetry in species richness between clades and how clade richness change through time are two of the most intriguing patterns in biodiversity. Those differences are ultimately regulated by different speciation and extinction dynamics which, respectively, adds and remove species across geologic time scales. Understanding how species richness change through time and their underlying processes are ultimate goals in macroevolutionary studies, but it is also important to acknowledge the biases inherent to the different types of data (fossils and molecular phylogenies) used in such studies (Foote, 2000; Quental & Marshal, 2010; Rabosky 2010; Silvestro *et al*, 2014a). Those range from the degree of completeness of the information, to our ability and practice to identify and describe new taxonomic entities (i.e. species according to the taxonomical practice). Therefore the taxonomic practice is expected to have a great influence in diversification dynamics inferences, and although some numerical simulations suggest the taxonomic patterns are congruent with natural processes (Sigwart *et al*, 2017), the impact of taxonomy in macroevolutionary studies should never be neglected (Rabosky, 2016).

Due to the timescale and nature of the macroevolutionary phenomena, the fossil record is the most direct source of data to study biodiversity patterns and diversification dynamics (Quental & Marshall, 2010; Marshall 2017). The usefulness of such kind of data strongly relies on the large compilation of fossil occurrences, either directly from fossil specimens deposited on museums or from the primary literature. This can be, depending on the taxonomic broadness, extremely time-consuming, but its use has been made easy and practical due to the development of fossil online databases, such as the Paleobiology Database (PBDB - <<u>http://paleobiodb.org/</u>>; <http://fossilworks.org/>) the New and Database and Old Worlds (NOW <http://www.helsinki.fi/science/now/>). Both databases are the result of non-governmental, nonprofit, scientific organizations which made fossil data occurrences from all the world available to virtually everyone with a computer and internet connection. Even though such databases represent an ongoing and active effort of the scientific community to maintain the best data always available, we should expect to find mistakes as in any large database (e.g. Genebank, or GBIF - Zizka et al, 2019). It is also possible that different scientists simply disagree with some of the information stored on those databases. Those mistakes/disagreements range from simple errors of digitation to different taxonomy practices, that could either be outdated or represent the preferred taxonomy of curators of a given study group.

In fact, the blind use of such large fossil databases has been the subject of harsh criticism from some experts which might recognize errors in some of their groups of interest, in particular an outdated taxonomic treatment (e.g. Stearn, 1999; Prothero, 2014). Another criticism is that many compilations are made and used by researchers that are not necessarily taxonomic experts (Stearn, 1999; Prothero, 2014). Applying different taxonomic decisions to a given dataset can lead to different patterns of diversity, but very few publications have directly assessed the actual effect of such potential errors to the estimation of diversity dynamics throught time. The only three examples, to our knowledge, of descriptions of the impact of adopting different taxonomics

are Maxwell & Benton (1990), Adrain & Westrop (2000), and Prothero (2014) papers. Some researchers do recognize these limitations on fossil databases and will perform taxonomical reviews or curation on the data prior to analysis (as, for example, in Cantalapiedra *et al*, 2015 or Pires *et al*, 2016). This is typically done in a framework where most changes are not directly identified and made available to the reader, which can considerably reduce the reevaluation of such decisions and the replicability of the results obtained. While Prothero's analysis of some mammal groups presents a visual description of the changes in biodiversity curves of some families of mammals under different taxonomic treatments, Adrain & Westrop (2000) found that more than 70% errors found in trilobites for the Sepkoski's database is in fact randomly distributed and does not result in a biased description of diversity dynamics. Benton and Maxwell (1990) also reports small changes between different taxonomic arrangements. It is important to note, however, that the data used in Maxwell & Benton (1990) and Adrain & Westrop (2000) refers to higher taxa (families and genera, respectively), while Prothero's (2014) investigation used data in the genus and species levels.

There is also a perception, backed up by good arguments, that an analysis at a higher taxonomic level should be more robust to the incompleteness of fossil record given that the fossil record is more complete as we move up in the taxonomic ladder. This has been used to advocate for higher level analysis which could in theory be used as a proxy for the species level dynamics. This practice called "taxonomic surrogacy" is often used to analyze paleontological data (e.g. Raup & Sepkoski, 1986; Lu *et al*, 2006; Jablonski, 2017) when species-level identification is impossible or when the record is too sparse at the species level dynamics?; (2) do differences in taxonomic treatments affect similarly diversification analises done at different taxonomic levels (e.g. families, genera, species)? Although it might seem intuitive that higher level taxonomic level analysed, it is expected that even a simple and quick review generates better data than no review at all, but it is an open question if simple curatorial work might sufice or if there is a continual substantial improvement as further taxonomic work is done.

The aim of this research is to evaluate the effect of different degrees of taxonomic curatorial work over diversification inferences of Ruminantia. The ruminants (sub-order Ruminantia, Class Mammalia) include nearly 200 extant (Cantalapiedra *et al*, 2014) and nearly 1200 extinct species, distributed in 6 extant and 13 extinct families over 55 million years of evolutionary history (Cantalapiedra *et al*, 2015). The fossil taxonomy of the group is knowingly influenced by the incorrect descriptions and unjustified identifications made by previous taxonomists (Prothero (2014) cites some of them). Only recently Frick's incorrect taxonomical practice has been revised (e.g. Prothero 2008; Vislobokova 2013), which means that most of the reclassification needed is still a work in progress (Prothero, 2014) and is in need to be incorporated in large databases (Prothero, 2014). Therefore, the Ruminantia group has a high species richness and a taxonomical and scientific context appropriate to give valuable evidence on our questions.

More specifically, we asked: (1) What proportion of the data entrances have typos, outdated taxonomy treatment, or represent species that, according to current taxonomy, should be synonimized?; (2) How different taxonomic reviews of the data change important data properties and patterns?: (3) How different degrees of taxonomic reviews shape the patterns found?; (4) Does the use of higher taxa render taxonomic differences at lower levels negligible?; (5) Are higher taxa a good proxy for species level dynamics? To answer these questions we compare the diversification dynamics inferences (temporal estimates of speciation and extinction rates) of Ruminantia (Mammalia) resulting from the data directly downloaded from NOW and PBDB databases after the implementation of different taxonomic reviews of this same initial dataset. This was done in a bayesian framework that explicitly accounts for temporal and among species variation in fossil preservation. All analisvs were done both at species and genus levels. We used two types of "raw" data (one simple processed and one not processed at all) and two taxonomic opinions: one based on expert reviews (Cantalapiedra et al, 2015), and a semiindependent review, made by the authors without the supervision of any taxon specialist and almost completely based on the primary literature (see details in Supplementary file 1). All revisions are made under an explicit framework (see material and methods), given that we also advocate for a more explicit way for others to reaccess and reevaluate what was done.

Material and Methods:

Data collection and curatorial work

Our framework is summarized in Supplementary Figure 1. We downloaded all occurrences from the Paleobiology Database (PBDB – <u>http://paleobiodb.org</u>) and the New and Old Worlds Database (NOW – Fortelius, M.: Coordinater, <u>http://www.helsinki.fi/science/now</u>) assigned to the "Artiodactyla" group. We then excluded non-ruminant families (Supplementary Table 1). One of the databases used, hereafter called "ALL dataset", has this treatment only. For the other two datasets, we excluded all occurrences in the database with explicit taxonomy uncertainty at the level of the analysis (i.e. occurrences where species or genus were marked with qualifiers such as "sp.", "cf.", "aff.", and "?", which codes for different degrees of taxonomic uncertainty - Bengtson, 1988 - see supplementary table 1).

As highlighted by Prothero (2014), in both NOW and PBDB databases the taxonomy of a given occurrence may be outdated. To understand the magnitude of the outdated taxonomy, we compared the species list given by PBDB and NOW with more than 310 primary references (Supplementary Table 2). We used only the following three types of primary literature: (a) Systematic/taxonomic review of the species, genus, family or other higher taxa (e.g. Prothero, 2008; Vislobokova, 2013), (b) Recent papers which described new species and compared them to other valid members of its group (e.g. Mennecart & Métais, 2015; Tabrum & Métais, 2007), and (c) Descriptions of paleo-communities (e.g. Vislobokova 2008; Liu *et al*, 2012), the later was used even if it does not represent a formal revision because it help inform if at least a given taxonomical name is used or is valid in the literature at a given time. The types of primary

literature, and the taxonomical practices they reflect, were used in a hierarchical order: if type (a) was not available, we used type (b) as a base for review, and if neither (a) or (b) was available, we used type (c). If there were conflicting opinions (e.g. some species is said to be valid in one article but invalid in other article) in the same type of literature, we used the year of publication and the author's name to make a decision. We gave preference for the taxonomic opinions stated by the authors of chapters from Prothero and Foss's (2007) book, as an attempt to keep consistency between taxonomic practices of different authors. Based on the same primary literature, we removed all occurrences which were later identified as non-ruminants (but were identified as ruminants in the databases). We also synonymized different names which were indicated in the primary literature as synonyms. Given that we did not have direct access to the specimens, and that we are not experts in Rumiantia, we did not split a given name into two or more species because this would not be possible with our current access to the data and expertise. We annotated each name we maintained, synonymized or excluded in the following categories: "typo" (synonymized), "synonymized name" (synonymized), "invalid name" (excluded), "no literature available" (maintained) and other minor categories (see supplementary file 1 to a detailed description of the criteria used and categories signaled). We compiled a full list of synonyms (supplementary table 2) and another list comparing all names in each dataset (Supplementary File 1, Supplementary Table 3). These supplementary lists have the full account of names, including the typos found in all databases. The fully revised dataset, hereafter called "OUR dataset" contains all data with the taxonomic synonyms mentioned above. A third dataset, hereafter called "RAW dataset" also has the taxonomical uncertainties removed, but does not have a synonymizing effort because its curation was completely conducted with automated scripts. To avoid including pseudo-replicates within and between PBDB and NOW, in both RAW and OUR datasets, we removed all species with the same early and late bounds of geologic time and geographical coordinates within a threshold of 0.01 decimal degrees. As a fourth database treatment, we downloaded the dataset used in Cantalapiedra et al (2015) from their supplementary material, and used it without any treatment, except regarding temporal resolution (hereafter, this dataset is named "CAN dataset"). Because data with low temporal resolution introduces too much uncertainty in the analysis we preferred to exclude any occurrence (of any dataset) with a temporal range larger than 15 Myr. The genera analysis where made simply removing the specific epithet for ALL, CAN and OUR datasets, but for RAW dataset we included back occurrences where the explicit taxonomic uncertainty was only at the species level (i.e. markers as "cf." and "sp." - please see R scripts annotated in the boxes of supplementary figure 1). The main difference between ALL and RAW datasets at the genus level is that RAW did not have names with uncertain identification at the genus or higher taxonomic levels.

Birth-death diversification analysis

To infer diversification dynamics and estimate its associated uncertainty we fitted our data with a hierarchical Bayesian framework (Silvestro *et al*, 2014a; Silvestro *et al*, 2014b;

Silvestro *et al*, 2018a). Here fossil occurrences are modeled as the result of two interacting processes: preservation and diversification rates, the latter decomposed into speciation (or origination for higher taxa such genera) and extinction rates. The method jointly estimates the preservation rate, the times of speciation/origination and extinction of each lineage, and the rates of extinction and speciation/origination through time. Each estimation is sampled under a birth-death reversible jump markov chain monte carlo (BDRJMCMC - hereafter birth-death model) framework (Silvestro *et al*, 2018a), which gives direct access to the uncertainty (i.e. posterior of parameter percentile intervals of the sampled distribution) associated with the data across geological time (Silvestro *et al*, 2014a; Silvestro *et al*, 2014b; Silvestro *et al*, 2018a).

Under this Birth-death model, preservation is modeled as a stochastic non-homogeneous Poisson process that is a function of time and the mean preservation rate q, which is estimated from data. There are three options to model the preservation: (1) mean preservation rate qparameter estimated from data, with preservation rate varying across taxa using a Gamma distribution model (mG model) with a shape parameter also estimated from data; (2) mean preservation rate (q) varying through time inside predetermined and specific time window intervals without preservation variation among species (qShift model); (3) preservation simultaneously varying across taxa (one shape parameter for all time windows estimated from data) and time (mean preservation rate q estimated from data inside specific time intervals) (mG + qShift model). For the options 2 and 3, where preservation varies in time, we used each epoch boundary according to GSA geologic time scale to delimitate the time intervals. This was justified by some properties of the data (see Supplementary File 2 for a detailed description) and the expectation that preservation might be better at specific time ranges or at the recent. For each dataset we ran 110000000 iterations, discarding the first 10000000 as burn-in, sampling every 100000 iterations to obtain posterior distributions for each parameter. The PyRate framework also needs a list of species within fossil data that are still living and the total number of extant species (including those who are not present in the fossil data). This information was obtained from Cantalapiedra et al (2014). To improve the convergence and sampling of our Birth-death model, we changed the default parameters of its integrator (see Supplementary File 3) using parameters suggested by Daniele Silvestro (personal communication). As fossil occurrences often are assigned to a temporal range and not a single age, we randomly drew ages within the range of each occurrence, generating 100 randomized replicas of each dataset before running the MCMC algorithm. We repeated all analysis in genera and species taxonomic levels, totalizing 2400 runs (4 datasets, 3 preservation models, 2 taxonomic levels, each with 100 replicas). Finally, to reduce the size of the final file for each dataset (after combining all replicates), we checked the ESS values of each replicate and pruned each MCMC replica to only 100 randomly drew iterations. For each preservation scenario and taxonomic level these MCMC sub-samples of all replicas were joint together to represent a single analysis. Effective sample sizes (ESS) were assessed using the function ESS in the LaplacesDemon R package (Hall, 2011). Although the comparison of ESS value for all analysis showed that the qShift preservation model is better

sampled, all models and most replicas had sufficient independent sampling (ESS values > 200) (see supplementary file 3).

Comparing different diversification inferences

The datasets analyzed resulted on the following estimates: i) rates through time, ii) the posteriors of all parameters of the birth-death model (including the preservation rate), iii) the posteriors for the times of origin and times of extinction for each lineage in each analysis. To compare datasets, we generated rates through time plots to summarize the outputs of PyRate for each combination of dataset, taxonomic level and preservation model. We assigned two rates as significantly different if there is no overlap between the 95% highest posterior density (HPD) intervals of their two estimates. The support for a given number of rate shifts and their time position (and hence how rates varied through time) can be assessed from the birth-death model sampling frequencies and their associated Bayes Factors (Silvestro et al, 2018a). To generate the diversity through time (DTT) curves we computed the number of living species/genera through time (0.1 My resolution) using only the longevities of sampled lineages, which does not fully deal with the imperfection of the fossil record (as a given species may have existed but leaved no fossil occurrence). To build this graph we considered the longevity of each lineage (species or genera, depending on the analysis) to be determined by the interval defined by the medians of the posterior distributions of the times of origin and the posterior distributions of the times of extinction for each lineage. To further compare the effect of different taxonomic treatments we compared the absolute values estimates for rates through time, the number of rate shifts, timing of those rate shifts, and number of species through time.

Results

Direct comparisons between the datasets

Each database has its own percentage of names with typos, invalid or synonymized names. But, as expected, using OUR dataset as reference, the number of species and occurrences with these tags decrease progressively from ALL (53.7% of names) to RAW (30.2% of names) to CAN (20.8% of names) databases (supplementary table 4). Even with these differences, we found a similar number of occurrences per lineage through time, both in magnitude and trajectory in time, when comparing species databases, except for the genus level analysis where the ALL dataset seems to have considerably more occurrences per lineage (table 1 and supplementary figure 2). Also note that if we compare total number of occurrences and total number of species or genus names then there is a great difference between databases, especially between ALL dataset and the others (see table 1). All databases have their lowest temporal resolution occurrences concentrated closer to the origin of the clade (before 40 Ma, (Supplementary figure 3, 4, 5, 6, and 7). The temporal resolution in CAN dataset is consistently

the smallest (i.e. most uncertain) of all datasets between 45 and 23 Ma, with a considerable lower resolution around the 35-22 Ma interval (supplementary figure 8 and 9). This low resolution is evidenced when it is divided by the richness of lineages (supplementary figure 8 and 10). However, the most evident difference between the datasets is their "crude" (sum of living species as a literal reading of the fossil record) richness through time (supplementary figure 9). The ALL dataset has a lot more species than any other dataset (table 1, supplementary figure 2). The CAN dataset has the smallest richness of all, but after 22 Ma it presents a richness very similar to OUR database (supplementary figure 10). In the genus level, ALL and RAW have similar richness through time, the same being true for the OUR and CAN, although the last two have considerably less lineages then ALL and RAW datasets (Supplementary figure 10).

Dataset	Number of occurrences	Number of genera	Number of species (names)	Resolution (my) before 23 Ma		Resolution (my) after 23 Ma			
				5% perc.	med.	95% perc.	5% perc.	med.	95% perc.
ALL	19218	497	2135	0.6	3.3	5.8	0.06	1.8	5.32
RAW	7992	454	1479	0.6	3.3	5.8	0.11	1.5	4.55
OUR	7841	411	1137	0.6	3.3	5.8	0.11	1.5	4.55
CAN	9234	390	1246	0.68	3.3	10.44	0.11	1.37	4.5

Table 1: Number of occurrences, genera names and species names for each dataset at the species level

Birth-death model analysis

There were few subtle differences between the different preservation models but in general the results are very similar among preservation models, both in species and genus levels (see supplementary figure 11, 12, 13 and 14). Given that the relevant differences are among different datasets and that the preservation model mG+qShift is the most complete model, we will focus only on the results of this model. The rates through time plots with ALL, RAW and OUR datasets, which are all derived from the same initial dataset (see comparison with CAN below) showed similar trends at the species level, except for a peak in speciation around 20 Ma for the ALL dataset and a drop in speciation rate for the same dataset from 30 to 20 Ma (figure 1, see also supplementary figure 11). The three datasets indicated a constant, although small, positive diversification from the rise of the Ruminantia clade until around 5 Ma, when all datasets suggested a negative diversification (figure 1).

In the species level, the timing of speciation rate shifts with low significance are roughly similar for ALL, RAW and OUR datasets between 32 Ma and 8 Ma, although not all time shifts were found in all datasets (figure 1). We also found one highly-significant speciation rate shift around 6 Ma which was detected in all the three datasets. The ALL dataset has two exclusive highly-significant speciation rate shifts around 21 Ma and 19 Ma, which creates a pulse in

speciation representing the most variable pattern of rate variation between the three datasets. The OUR dataset has one exclusively low-significant speciation rate shift around 1 Ma. Overall, there is overlap between the HPD in all three datasets, except between 19.2 Ma and 15.6 Ma for the ALL-RAW comparison and between 1.4 Ma and the present for the OUR-RAW comparison.

The extinction patterns in the species level look very similar for these three datasets: there are highly-significant extinction rate shifts around the same time between ALL, RAW and OUR datasets around 6 Ma, 4 Ma and 1 Ma. Two low-significance extinction rate shift occurred at similar timings between OUR and ALL dataset around 42 Ma and around 7 Ma, and at similar times between ALL and RAW datasets around 13 Ma. The ALL dataset has one exclusively low-significance extinction rate shift around 17 Ma, and one exclusively highly-significant around 3 Ma. The RAW dataset has one exclusively highly-significant extinction rate shift around 7 Ma and one exclusively low-significance extinction rate shift around 17 Ma, and one exclusively highly-significant extinction rate shift around 7 Ma and one exclusively low-significance extinction rate shift around 3 Ma. Overall, there is considerable overlap between the HPD in all three datasets, except between 1 Ma and the present for the ALL-OUR comparison, and between 20.9 Ma and 18.9 Ma for the ALL-RAW comparison.

The net diversification rates have patterns almost identical between the three datasets, with no overlap in HPD occurring only between 1 Ma and the present for the ALL-OUR dataset pair. The other striking difference in net diversification rates was the peak in rate for the ALL dataset around 20 My which was the result of a peak in speciation (figure 1). The lineages through time plots also showed a very similar temporal dynamics in the richness for all datasets, with RAW and OUR being almost indistinguishable. The ALL dataset had similar dynamics, but with higher numbers of species throughout the whole time, and a peak diversity that reached up to 500 coexisting species in the Pleistocene (compared to 300 in the RAW and OUR datasets). The results at the genus also showed great similarity between ALL, RAW and OUR datasets, with origination rate being slightly more variable, in particular between OUR dataset and the other two towards the present (supplementary figure 13). The genus level analysis also showed similar temporal patterns of rate shifts among datasets (one striking exception was the highly-significant shift in origination rate for OUR dataset around 1 Ma), and a complete overlap between the HPDs of the rates for the three datasets (supplementary figure 13) across all time.



Figure 1. Rates through time (RTT) plots with HPD = 95% for the posterior of speciation, extinction and Net Diversification rates for ALL (orange), RAW (red) and OUR (green) datasets. The bottom lines represent times of rate shift with low-significance, and the squares represent the times of highly significant rate shifts. Diversity through time (DTT) plots based on Ruminantia times of speciation and extinction estimated from PyRate for ALL (orange), RAW (red) and OUR (green) datasets. The grey and white vertical bars indicate the time interval within the preservation rate were constant (but between intervals the preservation rate could be different). The dashed line in the upper right panel delimitates when the Net Diversification equals zero.

The CAN dataset at the species level indicates very different diversification dynamics to the OUR dataset (figure 2) as well as to the other two (supplementary figure 37). It is particularly different before 20 Ma where the CAN dataset has 1 exclusively low-significance speciation rate shift around 43 Ma, and one exclusive highly-significant speciation rate shift around 15 Ma. More importantly, the CAN dataset suggested a much higher speciation rate up to around 21 My, where the speciation rate begins to drop until a highly-significant speciation rate shift around 15 Ma. During this initial period (between 41 Ma and 21 Ma) there is no overlap in speciation rate HPD between OUR and CAN datasets. The CAN dataset also had two exclusive lowsignificance extinction rate shift around 43 Ma and 20 Ma and one exclusive highly-significant extinction rate shift around 3 Ma. All the other rate shifts present in CAN dataset have correspondence with shifts found in OUR dataset. Although the HPD for the extinction rate is quite large for the CAN dataset, the initial extinction dynamics (up to around 20 Ma) seemed very different, with considerably higher rates for the CAN dataset when compared to OUR dataset. In fact, there is no overlap in extinction rate HPD between 38.4 Ma and 33.1 Ma for the OUR and CAN datasets. The Net Diversification rates have similar patterns between CAN and OUR datasets, with overlap in the HPD of the Net Diversification rate during all the clade history. This similarity probably exists because the differences found on speciation and

extinction rates canceled each other out here. The overall trends in the lineages through time plots are similar between the two datasets, especially after 20 Ma. Two relevant differences are that OUR dataset shows a considerably higher richness prior to 20 Ma, and the CAN dataset shows a much more variable richness through time (figure 2). Both differences agree with a high turn-over rate estimated for the CAN dataset prior to 20 Ma. After 23 Ma CAN dataset starts increasing its richness rapidly until CAN show a slightly greater richness than OUR between 10 Ma and 3 Ma. Both datasets have their highest richness in the Pleistocene.



Figure 2. Rates through time (RTT) plots with HPD = 95% for the posterior of speciation, extinction and Net Diversification rates for CAN (blue) and OUR (green) datasets. The bottom lines represent times of rate shift with low-significance, and the squares represent the times of highly significant rate shifts. Diversity through time (DTT) plots based on Ruminantia times of speciation and extinction estimated from PyRate for CAN (blue) and OUR (green) datasets. The grey and white vertical bars indicate the time interval within the preservation rate were constant (but between intervals the preservation rate could be different).

As ALL, RAW and OUR datasets presents similar patterns of diversification, we focus our comparisons between genera and species levels analysis only for OUR and CAN datasets (figure 3, but see suplementary figure 37). For OUR dataset, all origination rate shifts found in the genera level have a temporal correspondent with equivalent significant shifts in the speciation level, with one exception: the 1 Ma shift has high-significance in genera-level but only low-significance in the species level. Additionally, the species level analysis has a much larger number of speciation low-significance rate shifts. As expected the speciation rate shifts found in the species level analysis have a temporal dynamics is very similar. All the extinction rate shifts found in the species level analysis have a temporal correspondent in the extinction in the genus

level analysis, with one exception: a low-significance extinction rate shift around 41 Ma that is only present in the species level. The genus level analysis also have one exclusive extinction rate shift around 9 Ma. As expected the extinction rate at the species level is higher than extinction rate at the genus level but the temporal dynamics is very similar. The Net Diversification rates have close patterns between species and genera analysis for OUR dataset, with overlap in the HPD of the Net Diversification rate during all the clade history.



Figure 3. Rates through time (RTT) plots with HPD=95% for the posterior of speciation (or origination in the case of genera), extinction, and Net Diversification rates for CAN (blue) and OUR (green) datasets. The darker colors represents the genera-level analysis and the lighter colors represents the species-level analysis. The bottom lines represent times of rate shift with low-significance, and the squares represent the times of highly significant rate shifts. The grey and white vertical bars indicate the time interval within the preservation rate were constant (but between intervals the preservation rate could be different).

For CAN dataset, the differences between genera and species analysis are more remarkable (figure 3): only two shifts in origination and speciation rate have time-correspondence but one is highly-significant only in genus level analysis (around 20 Ma) and the

other is highly significant only in species level analysis (around 14 Ma). The other speciation/origination rate shifts have different time locations between genera and species-level analysis, and the species-level analysis has twice the number of rate shifts. As expected the speciation rate is higher than origination rates, but surprisingly the temporal dynamics is very different. The only highly-significant extinction rate shift found in CAN dataset in the genera level analysis has a temporal correspondent in the extinction at the species level. The other highly-significant extinction rate shifts found in the species level have time correspondence with low-significance extinction rate shifts at the genera level (around 6 and 5 Ma), with one exception: the highly-significant shift around 3 Ma, which has no correspondence in generalevel. All the other low-significance extinction rate shifts have no time correspondence when comparing the analyses on different taxonomic levels. The overall extinction dynamics is reasonably similar except for a peak in extinction rate found around 5 Ma only for the species level analysis. The net diversification dynamics suggests positive diversification up until around 8 Ma when the analysis at different taxonomic levels start to show marked differences (figure 3). Both CAN and OUR suggest a negative diversification close to the present (figure 3), which is mirrored by a drop in diversity in both species and genus level diversity through time plots (figures 2 and 4). At the genus level, the diversification temporal dynamics of origination and extinction (and hence net diversification) is almost identical between CAN and OUR datasets (figure 4), with only one major difference: OUR dataset has a burst in origination rates between 5 Ma and 2 Ma, but even during this period the HPDs of all rates overlap between CAN and OUR at the genus level.



Figure 4. Rates through time (RTT) plots with HPD = 95% for the posterior of origination, extinction and Net Diversification rates for CAN (blue) and OUR (green) datasets at the genus-level. The bottom lines represent times

of rate shift with low-significance, and the squares represent the times of highly significant rate shifts. Lineages through time (LTT) plots based on Ruminantia times of speciation and extinction estimated from PyRate for CAN (blue) and OUR (green) datasets. The grey and white vertical bars indicate the time interval within the preservation rate were constant (but between intervals the preservation rate could be different).

Another important result from PyRate is the description of how the preservation rate (measured in "number of occurrences per lineage per my" - Silvestro et al, 2014a) varied through time (figure 5). Our results show that, at the species level, the preservation rates increase through time, with consistently low preservation rates prior to 23 Ma (all medians of q posteriors smaller than 1.2 occurrences per lineage per my, only two of them greater than 1 occurrence per lineage per my). At the Miocene, preservation rates increase considerably (smaller median = 1.5occurrences per lineage per my, most medians greater than 2 occurrences per lineage per my). After 5.3 Ma, preservation rates practically double (smaller median = 2.2 occurrences per lineage per my; most medians greater than 6 occurrences per lineage per my). Even though all datasets show this overall temporal tendency of increased preservation towards the present, we found some variation between the datasets. The ALL dataset is the only one with a continuous increase in preservation rate as we approach the present (median during > 56 Ma interval = 0.9occurrences per lineage per my; median between 1.8 Ma and the present = 26.3 occurrences per lineage per my). The other datasets have more complex patterns, and preservation might either increase or decrease slightly between consecutive time periods (figure 5). Until 23 Ma, CAN dataset has the lowest preservation rate of all datasets, with a specially low preservation rate (median = 0.4 occurrences per lineage per my) during the 33.9 - 23.03 Ma interval. During the same time, RAW and OUR datasets have very close preservation rates. During the 23.03 - 15.97 Ma interval, the preservation rate in CAN increases 8 folds (median = 3.2), becoming 50% greater than OUR (median = 2.3) and RAW (median = 2.1), and very close to ALL preservation (median = 3.6). Between 15.97 Ma and 5.3 Ma, CAN have preservation rates close to OUR and RAW, and after 5.3 Ma CAN dataset has a much greater preservation rate than OUR and RAW datasets, with the largest differences in the interval between 1.8 Ma and the present (medians: CAN = 11.6; OUR = 7.2; RAW = 6.3).



Dataset 🖨 ALL 📫 CAN 📫 OUR 📫 RAW

Figure 5. Posterior distribution of the mean preservation rate (*q* parameter on PyRate) estimated for each time interval within analyses with mG+qShift preservation model. CAN dataset did not have occurrences in the first (before 56 Ma) interval. ALL dataset is not shown in the last two intervals (after 2.58 Ma) for visualization purposes (median ALL interval between 2.58 Ma and 1.8 Ma = 24.75 occurrences per lineage per million years; median ALL interval between 1.8 Ma and the present = 26.33 occurrences per lineage per million years).

Discussion

Our results suggested that using different preservation models did not strongly influenced the diversification description (Supplementary figures 11, 12 13 and 14). One possible explanation is that there are roughly few preservation regimes through the history of ruminants, a number of regimes sufficiently low to be accommodated by the 4 classes of preservation generated by the mG model. So the variation in preservation rate captured by the mG model may in fact accommodate temporal variation in preservation by allocating species in the same time period the same preservation regime described by one of the 4 classes of preservation of the mG model.

As intuitively expected, our results showed that curatorial work on large databases can influence subsequent macroevolutionary analysis. Perhaps not as intuitive, this effect is far from linear with respect to the amount of effort put in, and several potential mistakes in the data are in fact directly documented in the database itself. Interestingly, our analysis suggested that the speciation dynamics showed more differences between datasets than extinction dynamics which was quite similar, at least in terms of trends, between different datasets. Diversification rates, being the difference of those two rates, showed a somewhat intermediate level of difference between the datasets.

When looking at the comparison where the only difference is the amount of curatorial work (ALL, RAW and OUR datasets all start with the same original data), we see that, as expected, the macroevolutionary patterns for the least curated dataset (ALL dataset) showed considerable differences in both speciation and extinction dynamics when compared to the dataset curated using more resources (OUR dataset). This result resonates the common sense and previous conclusions (Prothero, 2014) that using datasets straight from the databases might lead to bias or artifacts. One prominent example was the speciation pulse surrounded by two highlysignificant shifts which we found around 21 Ma for the ALL dataset. Additionally, diversity through time estimates showed considerable differences (but given that the PyRate framework does not explicitly accounts for unobserved taxa in the fossil record, the diversity comparison between datasets must be seen with caution). The ALL dataset had, during almost all the history of the clade, the highest number of species and the difference from the other datasets tends to stay proportionally the same until the diversity peak, which has almost 60% more species than the other datasets. That said, the robustness of rate estimates has been thoroughly tested and PyRate has been shown to accurately estimate speciation and extinction rates, even under very high levels of incompleteness of the fossil record (Silvestro et al, 2014a; Silvestro et al, 2014b; Silvestro et al, 2018a). Hence the rate dynamics comparison is likely to be robust and it suggested that some of the potential anomalies (e.g. peak in speciation rate for ALL dataset) could have been interpreted as real phenomena if no curatorial work had been done.

Perhaps less expected, a minimum level of curatorial work, here represented by the RAW dataset, already removes most of those anomalies and differences seen between the dataset with no curatorial work (ALL dataset) and the one with most curatorial work (OUR dataset). It is important to note that this minimum curatorial work consisted only of the removal of uncertainties already indicated by the database itself (supplementary table 1) with automated scripts. For example, the main difference between the ALL dataset and the RAW dataset was the removal of 11226 occurrences which the species identity was not certain (supplementary table 4), as well as very few geographical replicates which occurred within each database (see supplementary figure 1). The removal of those occurrences resulted on the removal of 656 species from the ALL dataset to produce the RAW dataset with 1479 species (see supplementary table 4).

The RAW dataset resulted in essentially the same macroevolutionary pattern than the one found in OUR dataset. This may implicate that, at least for ruminants, simple code-oriented data curation has the capacity to efficiently remove the strongest biases and artifacts present in crude data. After the first curatorial work, there are only 149 occurrences and 342 species more in the RAW dataset than in the OUR dataset which has a total of 7841 occurrences and 1137 species. The number of occurrences lost is smaller than the number of species because many species were synonymized to other species in the dataset. It is also worth mentioning that our taxonomic curatorial work per passed for more than a year of primary literature search, reading and data interpretation, in contrast with the few days of data-oriented curation (and instantaneous implementation) made to generate the RAW dataset. By no means we argue against a detailed

taxonomic curatorial work but is it interesting to note that these large databases already have information about data quality that if taken into account already strongly reduces some potential biases of large datasets. So in some sense we could say that the taxonomic mistakes/different opinions within those databases behave like "white noise" and not a bias *per se*. This is consonant to the pattern discussed previously by Adrain & Westrop (2000), which found that the errors in another large biodiversity dataset (the Sepkoski's database), even representing substantial amount of the data, were in fact distributed randomly and did not influenced the macroevolutionary description of the data. It is also worth mentioning that we suspect that our dataset was robust to those changes because it was a large dataset, and thus it is required a bias of large scale to have an influence over the diversification dynamics.

At first our results might seem to contradict Prothero (2014) who had previously pointed out that many groups within the Ruminantia should probably represent cases where taxonomical reviews were more justifiable and should generate major changes in the description of the diversity dynamics, but caution is necessary. First, it is likely that many of Prothero's necessary changes were already implemented in the databases prior to the day we downloaded the data. Second and more importantly, although 342 species (the difference between RAW and OUR datasets) might seem a high number at first, one has to remember that the rates estimated here are per lineage per million years, hence given the high diversity of Rumminants (about 1137 in OUR dataset) for a big effect to play out one would need a lot more species to be changed. This means that taxonomical reviews will have a stronger effect when the richness of the clades of interest is low, or when numerous taxonomic changes are concentrated at times where the clade of interest presents low richness such as the time of its emergence, periods prior to the clade extinction, or the moments where a lineage passed through diversity bottlenecks. Considering our results this might in fact explain the observation that the majority of rate differences between the four datasets, including the major differences between CAN (see below) and the other datasets, are in fact located at early Ruminantia diversification (before 20 Ma), a period of low diversity (see figure 2). Lastly, Prothero compared the dynamics simply by plotting diversity through time, without any statistical attempt to discern small variations to general trends. In fact it is worth noting that the Ruminatia sub-clades evaluated by Prothero (2014), show very similar general trends in diversity even though there are some small differences in absolute values and some small fluctuations, which of course, depending on the researcher question are important differences. This emphasizes another important aspect of those comparisons. It seems that comparing diversity through time might lead one to overemphasize those small differences, while fitting an underlying stochastic model to compare the dynamics (as we did here) might perceive those small differences as stochastic variation of a similar underlying birth-death model. This is akin to the reason why PyRate is so successful to massive under-sampling in the record as reported by several papers (e.g. Silvestro et al, 2014a; Silvestro et al, 2014b; Silvestro et al, 2018a).

To further investigate the effect of curatorial work down the line of macroevolutionary analysis we then compare OUR dataset to the CAN dataset, representing two different taxonomic

treatments. Similar to the results discussed above, speciation rate also showed more differences than extinction rate. In fact those represent the biggest differences in our comparisons, and although not easily explained, we suspect that the truncation on CAN dataset (Cantalapiedra *et al*, 2015) might have some effect on producing this difference. By truncating the data, we suspect that "edge effects" (Foote, 2000) were introduced, in this case by artificially synchronizing the origination of many lineages to produce a higher than expected speciation (origination) rate. Although this seems possible, it probably does not explain all differences because speciation rates remain high for the CAN dataset for quite a while after their initial time. Also, extinction rates, which would not be affected in the same way at the beginning of the clade history, show considerable differences during the same period of time (Eocene and Oligocene). It is possible that extinction differences result as a byproduct of speciation differences given that with such high species values and such a small diversity could only be produced with an also high extinction rate (see also argument below).

Those differences could in theory result from additional data not present neither in PBDB nor in the NOW database. In fact, CAN dataset have some exclusive species added by the authors in Cantalapiedra *et al* (2015), but the 65 species added represent a small fraction (5.2%) of the dataset, and seem to be concentrated in periods of time where the dynamics of CAN is very similar to OUR, as after 20 Ma (Supplementary figure 15). Therefore it seems that these species are insufficient to understand the major differences in the diversification patterns between OUR and CAN datasets.

Apart from data truncation, another clear difference between OUR and CAN datasets is the preservation rate, which is higher for CAN dataset after 23My but smaller before that. Before 23.03 Ma, but specially between 33 Ma and 23.03 Ma, CAN dataset has the lowest preservation. A very low preservation rate may raise the probability that PyRate accounts for many species unsampled in the fossil record. These many species raises the estimated speciation rate, which consequently raises the estimated extinction rate (as these unsampled species, under the framework estimative, "should die" before the preservation raises in the next time interval). As speciation and extinction are very close, the net diversification will be relatively low, which explains why the richness through time and the net diversification of OUR and CAN are much more similar than the speciation and extinction rates. Even though diversification rates might seem similar, the underlying macroevolutionary dynamics is considerably different. The results from the CAN data set suggest a very high turn-over rate that lasted for about 20 million years, while OUR data set suggest a turn-over rate with a considerably smaller magnitude during the same time period. Additionally, we suspect this diversification difference has more to do with the lower resolution of the data used by CAN and the fact that its dataset is truncated at the past which can cause edge-effects than that the use of different taxonomic practices.

Although one might intuitively think that the genus and species diversification dynamics are somewhat connected because species are nested within a given genus, and paleontologists have in fact used genus level as a surrogate for species level analysis, this is an empirical open question. Our results show that for most datasets, except for CAN, the general
macroevolutionary dynamics are essentially the same if one is interested in the general trends. These datasets, when analyzed under species or genera level have very similar temporal location of rate shifts and the same general direction of the rate change (i.e. increase or decrease). Our results show that, at least for ruminants, macroevolutionary analyses using the genera level are good proxies for the dynamics (i.e. temporal rate change) under the species level, at least for OUR dataset, potentially justifying the "taxonomic surrogacy" practice for the group. On the other hand, the opposite conclusion would be drawn when looking at the CAN dataset. Our results showed that the genus level analysis, although showing some superficial similarities with the species level analysis for the diversification dynamics, showed marked differences for speciation (origination) and extinction rates. We note though that this different descriptions of the dynamics between genus and species level, might in fact be related to the low preservation at the species level discussed earlier, nonetheless, if this was the preferred taxonomy the practice of "taxonomic surrogacy" would not be justified.

In general, our results indicate public databases might be less prone to produce biased results than previously thought. The decisions and changes of a specific taxonomic practice seems to act in analogous fashion to randomly-distributed errors, which is known to have little influence over the description of the dynamics (Raup, 1991). In our view, this pattern is a direct result of the great effort of the community to enter and maintain the best data available for paleobiologists around the world. Even though this is the first time this type of question is specifically addressed for the public databases in question, we are not the first to found that corrections in large databases can produce little change in the description of the diversity dynamics. Adrain & Westrop (2000) found that even that the classic Sepkoski genera database has its majority of data somewhat corrupted, the correction of all the errors and the increment of new findings did not result in a divergent description of change in diversity through time. Maxwell and Benton (1990) also found a similar pattern for different compilations of tetrapod families along the history of paleontology. Together, these results point in the direction that the reason taxonomical changes do not seem to strongly bias diversity dynamics descriptions may be that the amount of data in these datasets is so massive that it is needed a massive amount of taxonomical changes to influence the diversity descriptions. Although all the three datasets (Sepkoski's database, Vertebrate database of Maxwell and Benton (1990) and the database presented here) passed through major reviews on data, it seems that the scale of these changes is still not enough to change our general view on diversity. By no means we argue that this is evidence that taxonomic work is not relevant. Quite the opposite, it in fact suggests that huge amount of taxonomic work that has already been put into building those databases, including the possibility to directly inform uncertainties that can be easily removed with automated protocols, seem to sufficiently deal with the problem.

The initial taxonomical treatment that can be made exclusively using a simple code, has yet another interesting feature. It makes the review on data completely reproducible and transparent, as long as the investigators provide the script. In fact we argue that this practice to be followed even if the taxonomic treatments involve further taxonomic decision such as synonymizing two or more names. We also recommend the use of flow graphs, such as supplementary figure 1, to inform the reader about the magnitude of influence that each step of the review had on data. In the cases were a profound review is made (as in OUR dataset), we recommend the reporting of a table with all or at least most of the decisions made on data, as the supplementary table 2. This practice not only permits the scrutiny of data by other investigators, taxonomist or not, but also permits that different taxonomical practices, decisions or even different concepts of species could be analyzed and made explicit to the community more easily. In fact important arguments have been made for carefully using large databases (e.g. Prothero 2014), but most papers using PBDB and NOW database (including some done by the senior author of the current paper) do not report such decisions in the same level of detail we do here. This does not mean those works have not curated their data, but simply fail to make their decisions more transparent. As transparency and reproducibility are a major concern in current science, we argue that the use and sharing of scripts, functions or code in general seems convenient means to implement such protocols in paleobiology. We hope our results may encourage a more wider, transparent and reproducible view on the use of public paleobiological databases.

Acknowledgements: We thank Daniele Silvestro for help with the MCMC integrator and Mark Uhen for help with the use of the PBDB information. We also thank Juan Cantalapiedra for providing arguments over the taxonomy of some occurrences, and Paulo Guimarães Jr. and Eduardo Santos provided important suggestions at different stages of this study. M.L.J. was funded by Conselho Nacional de Desenvolvimento Científico (CNPq), Brasil (process number 130126/2017-9.). T.B.Q. would also like to thank FAPESP for financial support (grants # 2012/04072-3 and #2018/05462-6).

Competing interests: The authors declare no competing interests

Code availability: The scripts used can be downloaded at the authors github: https://github.com/mjanuario/Taxonomy_Januario_Quental_sup_material/tree/master/scripts

References

Adrain, J. M. and S. R. Westrop 2000: An Empirical Assessment of Taxic Paleobiology. Science 289:110-112.

Bengtson, P. 1988: Open nomenclature. Palaeontology 31:223-227.

- Cantalapiedra, J. L., M. Hernández Fernández, B. Azanza, and J. Morales. 2015: Congruent phylogenetic and fossil signatures of mammalian diversification dynamics driven by Tertiary abiotic change. Evolution 69:2941–2953.
- Cantalapiedra, J. L., R. G. FitzJohn, T. S. Kuhn, M. H. Fernández, D. DeMiguel, B. Azanza, J. Morales, and A. Mooers. 2014: Dietary innovations spurred the diversification of ruminants during the Caenozoic. Proceedings of the Royal Society B: Biological Sciences 281.
- Foote, M. 2000: Origination and extinction components of taxonomic diversity: general problems. Paleobiology 26:74-102
- Hall, B. 2011. LaplacesDemon: An R Software for Bayesian Inference. R package version 11
- Jablonski, D. 2017: Approaches to Macroevolution: 1. General Concepts and Origin of Variation. Evolutionary Biology 44:427–450.
- Liu, P., C. Deng, S. Li, S. Cai, H. Cheng, Y. Baoyin, Q. Wei, and R. Zhu. 2012: Magnetostratigraphic dating of the Xiashagou Fauna and implication for sequencing the mammalian faunas in the Nihewan Basin, North China. Palaeogeography, Palaeoclimatology, Palaeoecology 315–316:75–85.
- Lu, P. J., M. Yogo, and C. R. Marshall. 2006: Phanerozoic marine biodiversity dynamics in light of the incompleteness of the fossil record. Proceedings of the National Academy of Sciences 103:2736– 2739.
- Marshall, C. R. 2017: Five palaeobiological laws needed to understand the evolution of the living biota. Nature Ecology and Evolution 1:1–6.
- Maxwell, W. D., M. J. Benton, and N. Summer. 2007: Historical Tests of the Absolute Completeness of the Fossil Record of Tetrapods Historical tests of the absolute completeness of the fossil record of tetrapods. 16:322–335.
- Mennecart, B., and G. Métais. 2015: Mosaicomeryx gen. nov., a ruminant mammal from the Oligocene of Europe and the significance of "gelocids." Journal of Systematic Palaeontology 13:581–600.
- Pires, M. M., D. Silvestro, and T. B. Quental. 2015: Continental faunal exchange and the asymmetrical radiation of carnivores. Proceedings of the Royal Society B: Biological Sciences 282.
- Prothero, D. R. 2008: Systematics of the musk deer (Artiodactyla: Moschidae: Blastomerycinae) from the Miocene of North America. New Mexico Museum of Natural History and Science Bulletin 44:207– 223.
- Prothero, D. R. 2014: Garbage in, garbage out: The effects of immature taxonomy on database compilations of North American fossil mammals. New Mexico Museum of Natural History and Science Bulletin 68:257–264.

- Quental, T. B., and C. R. Marshall. 2010: Diversity dynamics: Molecular phylogenies need the fossil record. Trends in Ecology & Evolution 25:434–441.
- Rabosky, D. L. 2010. Extinction rates should not be estimated from molecular phylogenies. *Evolution: International Journal of Organic Evolution* 64:1816-1824.
- ———. 2016: Reproductive isolation and the causes of speciation rate variation in nature. Biological Journal of the Linnean Society 118:13–25.
- Raup, D. M., and J. John Sepkoski. 1986: Periodic extinction of families and genera. Science 231:833– 836.
- Raup, D. M. 1991: The future of analytical paleobiology. Pp 207-216. In: Gilinsky, N. L. and Signor, P. W., eds. Short courses in paleobiology. Paleontological Society: Knoxville. TN.
- Sigwart, J. D., M. D. Sutton, and K. D. Bennett. 2018: How big is a genus? Towards a nomothetic systematics. Zoological Journal of the Linnean Society 183:237–252.
- Silvestro, D., N. Salamin, and J. Schnitzler. 2014a: PyRate: A new program to estimate speciation and extinction rates from incomplete fossil data. Methods in Ecology and Evolution 5:1126–1131.
- Silvestro, D., J. Schnitzler, L. H. Liow, A. Antonelli, and N. Salamin. 2014b: Bayesian estimation of speciation and extinction from incomplete fossil occurrence data. Systematic Biology 63:349–367.
- Silvestro, D., A. Antonelli, N. Salamin, and X. Meyer. 2018a: Improved estimation of macroevolutionary rates from fossil data using a Bayesian framework. BioRxiv 18:316992.
- Silvestro, D., R. C. M. Warnock, A. Gavryushkina, and T. Stadler. 2018b: Closing the gap between palaeontological and neontological speciation and extinction rate estimates. Nature Communications 9.
- Stearn, C. W. 1998: Easy access to doubtful taxonomic decisions. Palaeontologia Electronica 2:1-4.
- Tabrum, A. R., and G. Métais. 2008: Pipestoneia douglassi, A New Genus and Species of Selenodont Artiodactyl from the Pipestone Springs Area, Jefferson County, Montana. Bulletin of Carnegie Museum of Natural History 39:83–96.
- Vislobokova, I. A. 2008: Main stages in evolution of Artiodactyla communities from the Pliocene-Early Middle Pleistocene of northern Eurasia: Part 2. Paleontological Journal 42:414–424.
- ——. 2013: Morphology, taxonomy, and phylogeny of megacerines (Megacerini, Cervidae, Artiodactyla). Paleontological Journal 47:833–950.
- Zizka, A., D. Silvestro, T. Andermann, J. Azevedo, C. Duarte Ritter, D. Edler, H. Farooq, A. Herdean, M.

Ariza, R. Scharn, S. Svantesson, N. Wengström, V. Zizka, and A. Antonelli. 2019: CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. Methods in Ecology and Evolution.

Supplementary material



Supplementary figures (printed version)

Supplementary figure 1. Data handling, curatorial guide and dataset delimitation. Numbers inside parentheses refer to genera-level analysis, when it differs from species-level analysis.



Supplementary figure 2. Occurrences per richness through time for CAN (blue) OUR (green) RAW (red) and ALL (orange) occurrences. Dashed lines represent genus-level dataset, continuous lines represent species-level datasets. The gray rectangles represent each qShift interval used in qShift and mG+qShift analyses.



Supplementary figure 3. Mean of occurrence timespan (duration between early and late bounds) through time for CAN (blue) OUR (green) RAW (red) and ALL (orange) occurrences. Dashed lines represent genus-level dataset, continuous lines represent species-level datasets. Note that as ALL, OUR and CAN datasets have the exact same occurrences in the genera and species level, the dashed and continuous lines are superimposed for these datasets. The gray rectangles represent each qShift interval used in qShift and mG+qShift analyses.



Supplementary figure 4. Occurrence timespan (duration between early and late bounds) distribution through time for ALL dataset occurrences. Species and genera dataset are equal. Lines represent the mean (continuous for species, dashed for genus level analises), and the hatch represents the 95% percentile. Note that as ALL dataset have the exact same occurrences in the genera and species level, the dashed and continuous lines are superimposed. The gray rectangles represent each qShift interval used in qShift and mG+qShift analyses.



Supplementary figure 5. Occurrence timespan (duration between early and late bounds) distribution through time for RAW dataset occurrences. Dark color represent the genera level dataset, and the lighter color represent the species level dataset. Lines represent the mean (continuous for species, dashed for genus level analises), hatches represent the 95% percentile. The gray rectangles represent each qShift interval used in qShift and mG+qShift analyses.



Supplementary figure 6. Occurrence timespan (duration between early and late bounds) distribution through time for OUR dataset occurrences. Species and genera dataset are equal. Lines represent the mean (continuous for species, dashed for genus level analises), and the hatch represents the 95% percentile. Note that as OUR dataset have the exact same occurrences in the genera and species level, the dashed and continuous lines are superimposed. The gray rectangles represent each qShift interval used in qShift and mG+qShift analyses.



Supplementary figure 7. Occurrence timespan (duration between early and late bounds) distribution through time for CAN dataset occurrences. Species and genera dataset are equal. Lines represent the mean (continuous for species, dashed for genus level analises), and the hatch represents the 95% percentile. Note that as CAN dataset have the exact same occurrences in the genera and species level, the dashed and continuous lines are superimposed. The gray rectangles represent each qShift interval used in qShift and mG+qShift analyses.



Supplementary figure 8. Occurrences mean timespan (time between early and late bounds) per species through time for CAN (blue) OUR (green) RAW (red) and ALL (orange) occurrences. Dashed lines represent genus-level dataset, continuous lines represent species-level datasets. The gray rectangles represent each qShift interval used in qShift and mG+qShift analyses.



Supplementary figure 9. Distribution of timespan (time between early and late bounds) of occurrences for CAN (blue) and OUR (green) datasets. The gray and white vertical rectangles represent different preservation time intervals used in qShift and mG+qShift analyses.



Supplementary figure 10. Richness through time for CAN (blue) OUR (green) RAW (red) and ALL (orange) occurrences. Dashed lines represent genus-level dataset, continuous lines represent species-level datasets. The gray and white vertical rectangles represent different preservation time intervals used in qShift and mG+qShift analyses. Panel A reflects the graph in full scale and Panel B is the same curve but with restricted axis for visualization purposes



Supplementary figure 11. Rates through time plots with HPD 95% for the posterior of speciation, extinction and net diversification rates for ALL (orange), RAW (red) and OUR (green) datasets in the species level analysis. The continuous line represents the median of the posterior distribution in each interval. The bottom lines represent low-significance times of rate shift, and the squares represent the times of highly significant rate shifts. The gray and white vertical rectangles represent different preservation time intervals used in analysis qShift and mG + qShift.



Supplementary figure 12. Rates through time (RTT) plots with HPD 95% for the posterior of speciation, extinction and net diversification rates for CAN (blue) and OUR (green) datasets in the species level analysis. The continuous line represents the median of the posterior distribution in each interval. The bottom lines represent low-significance times of rate shift, and the squares represent the times of highly significant rate shifts. The gray and white vertical rectangles represent different preservation time intervals used in analysis qShift and mG + qShift.



Supplementary figure 13. Rates through time plots with HPD 95% for the posterior of origination, extinction and net diversification rates for ALL (orange), RAW (red) and OUR (green) datasets in the genera level analysis. The continuous line represents the median of the posterior distribution in each interval. The bottom lines represent low-significance times of rate shift, and the squares represent the times of highly significant rate shifts. The gray and white vertical rectangles represent different preservation time intervals used in analysis qShift and mG + qShift.



Supplementary figure 14. Rates through time (RTT) plots with HPD 95% for the posterior of origination, extinction and net diversification rates for CAN (blue) and OUR (green) datasets in the genera level analysis. The continuous line represents the median of the posterior distribution in each interval. The bottom lines represent low-significance times of rate shift, and the squares represent the times of highly significant rate shifts. The gray and white vertical rectangles represent different preservation time intervals used in analysis qShift and mG + qShift.



Supplementary figure 15. The CAN dataset. The horizontal lines indicate the timespan of each occurrence in CAN dataset. The red horizontal lines indicate the occurrences of species which were added by Cantalapiedra et al (2015), and so are absent in the other databases. Those make up to 212 (2.3%) of occurrences. The light gray bars indicate the other categories of occurrences present in CAN dataset. The gray and white vertical rectangles represent different preservation time intervals used in analysis qShift and mG + qShift.



Supplementary figure 37. Rates through time plots with HPD 95% for the posterior of speciation, extinction and net diversification rates for ALL (orange), RAW (red) and CAN (blue) datasets in the species level analysis. The continuous line represents the median of the posterior distribution in each interval. The bottom lines represent low-significance times of rate shift, and the squares represent the times of highly significant rate shifts. The gray and white vertical rectangles represent different preservation time intervals used in analysis qShift and mG + qShift.

Supplementary file 1

from "Ruminatia macroevolutionary patterns are robust to distinct taxonomic treatments on data from paleontological online databases"

Detailed descriptions on the taxonomic reviews

Taxonomy reviews:

As highlighted by Prothero (2014), in both NOW and PBDB databases the taxonomy of a given occurrence may be outdated. To understand the magnitude of these outdated taxonomy, we compared the species list given by PBDB and NOW with more than 310 primary references (Supplementary table 2). In our final taxonomic decisions, we used only the following types of primary literature:

(A) Systematic/taxonomic review of the species, genus, family or other higher taxa (e.g. Prothero, 2008; Vislobokova, 2013)

(B) Recent descriptions of new species which compared the focal new species with other valid names in the group (e.g. Mennecart & Métais, 2015; Tabrum & Métais, 2007)

(C) Descriptions of paleo-communities (e.g. Vislobokova 2008; Liu *et al*, 2012)

Later on, we emailed Juan Lopez Cantalapiedra asking for advice only in species we didn't found conclusive primary references. He sent us tables which he used in Cantalapiedra *et al*, (2015) work to make his synonimizations. These tables contained the considerations of a variety of Ruminantia specialists as Dimitrios Kostopoulos, Alan Gentry, Gertrud Rössner, Pasquale Raia and others. Many of these researchers are the authors of many primary references we found and used before emailing Cantalapiedra. We read all the arguments and most references cited and took an informed decision about each name. When we used the content of this table, rather than the articles cited by the researchers, we assigned "Based on J. L. Cantalapiedra advising. Please check Supplementary file 1" as the references cited in the table J. L. Cantalapiedra sent us, we simply assigned the reference directly in the supplementary table 2.

Based on the references described above, we separated all the names of all datasets in four categories:

Valid name

Names were marked "valid" if we found evidence (see above) that it is currently in use according to the three types of primary literature described above. In other words, it was assigned as "valid" because it had a taxonomy article that concluded that this is the name for a given specimen or group of specimens (directly by type A of reference, indirectly by type B of reference), or simply because this name is used in articles other than taxonomic articles (type C or primary references).

Synonymized

We marked as "Synonymized" the names which we concluded must be synonymized to a "valid name". To do this we followed a given reference (see above and Supplementary Table 2). Hence in most curated data, the OUR database, we synonymized all occurrences under a "synonymized name" to a unique and specific "valid name".

No reference found

When, following the procedures above, we found no article or consideration about a name, we listed the name as "No reference found". We decided to include "no reference found" names in OUR dataset, as we did not found evidence that these names were invalid. We didn't made sensibility analysis to these data because there are few names listed as "no reference found" (please see supplementary table 4) and we think the influence of these occurrences would be too small to affect our conclusions.

Invalid

We listed as "invalid" and later removed from OUR dataset all occurrences of names that were explicitly said to be invalid (e.g. Grubb, 2000)

Туро

When a given name was (in our consideration) a clear typo of a valid named, we listed this name as "typo".

Accepted by cantalapiedra

When we finished our synonymizations, there were names which we listed as "synonymized", but were on Cantalapiedra *et al* (2015) final dataset. In these case, we labeled these occurrences as "accepted by cantalapiedra" to make explicit that those different databases dealt differently with these names.

Added by cantalapiedra

As said in Cantalapiedra *et al* (2015) article, the authors included in their (CAN) dataset occurrences which were not listed in PBDB and NOW at the time they downloaded the data. We listed as "added by cantalapiedra" the names which were in CAN dataset but not in any other, which gives us indirect evidence that these names were added by the authors of Cantalapiedra *et al*, (2015).

Non-Ruminant (after revision)

Some species are listed in the databases vaguely as members of the Artyodactyla group (and have no specification of, for example, the family of that species) or directly as members of one of the Artyodactyla families, but during the review we found evidence that they are not. In other words, there are species incorrectly signaled as possible ruminants, but we only have access to evidence of this during the taxonomical review, which means that these species are included in ALL and RAW databases. We labeled these species/occurrences as "Non-ruminant (after review)" and removed them from OUR database. These represent only a really small fraction of the data (supplementary table 4)

Removed with code

These are occurrences that we removed with code while transforming ALL in RAW dataset. The only dataset which they appear is ALL.

OBS: Note that after the beginning of the taxonomic revisions we decided to be more conservative in the data cleaning with code (while transforming ALL in Raw dataset), so some categories of names are not necessarily mutually exclusive. Also, if a given name is accepted or added by Cantalapiedra *et al* (2015) it could also fall in more than one of the categories listed above.

References

- Cantalapiedra, J. L., M. Hernández Fernández, B. Azanza, and J. Morales. 2015: Congruent phylogenetic and fossil signatures of mammalian diversification dynamics driven by Tertiary abiotic change. Evolution 69:2941–2953
- Grubb, P. 2000: Valid and invalid nomenclature of living and fossil deer, Cervidae. Acta Theriologica 45:289–307.
- Liu, P., C. Deng, S. Li, S. Cai, H. Cheng, Y. Baoyin, Q. Wei, and R. Zhu. 2012: Magnetostratigraphic dating of the Xiashagou Fauna and implication for sequencing the mammalian faunas in the Nihewan Basin, North China. Palaeogeography, Palaeoclimatology, Palaeoecology 315–316:75–85.
- Mennecart, B., and G. Métais. 2015: Mosaicomeryx gen. nov., a ruminant mammal from the Oligocene of Europe and the significance of "gelocids." Journal of Systematic Palaeontology 13:581–600.
- Pires, M. M., D. Silvestro, and T. B. Quental. 2015: Continental faunal exchange and the asymmetrical radiation of carnivores. Proceedings of the Royal Society B: Biological Sciences 282.
- Prothero, D. R. 2008: Systematics of the musk deer (Artiodactyla: Moschidae: Blastomerycinae) from the Miocene of North America. New Mexico Museum of Natural History and Science Bulletin 44:207–223.
- Raup, D. M., and J. John Sepkoski. 1986: Periodic extinction of families and genera. Science 231:833–836.
- Tabrum, A. R., and G. Métais. 2008: Pipestoneia douglassi, A New Genus and Species of Selenodont Artiodactyl from the Pipestone Springs Area, Jefferson County, Montana. Bulletin of Carnegie Museum of Natural History 39:83–96.
- Vislobokova, I. A. 2008: Main stages in evolution of Artiodactyla communities from the Pliocene-Early Middle Pleistocene of northern Eurasia: Part 2. Paleontological Journal 42:414–424.
 - —. 2013: Morphology, taxonomy, and phylogeny of megacerines (Megacerini, Cervidae, Artiodactyla). Paleontological Journal 47:833–950.

Supplementary file 2

from "Ruminatia macroevolutionary patterns are robust to distinct taxonomic treatments on data from paleontological online databases"

Determining the preservation models and the intervals with different preservation potentials:

In the PyRate framework, there is an option to allow different species to have different preservation potentials. This is done by, instead of assigning a single preservation rate to all species, using a model of preservation where the preservation rate varies among species according to a discrete Gamma distribution with shape (alpha), mean (q) and number of discrete categories used to approximate to the gamma distribution (k). The alpha and q parameters are estimated from the data depending on the preservation model used (see below), and the k parameter is specified by the user. We maintained constant k = 4 (default in PyRate).

We used this gamma preservation model (mG) in combination or not with another preservation model called "qShift", which allows the preservation process to be different in different, windows of time predetermined by the user.

Hence we ended up testing with the following preservation models:

mG: alpha and q are estimated from the data using the whole time series. This presumes no temporal variation in preservation rate, although in theory some of that temporal variation could be captured by different categories (k) given that no lineage lives forever and lineages with similar preservation potential could co-occur in time.

qShift: different q values are estimated from the data , one for each user-predetermined temporal interval. There is no preservation variation among species.

mG + qShift: different q are estimated from the data for each user-predetermined temporal interval. *Alpha* is constant for the whole time series but is estimated from the data.

To determine the temporal preservation intervals for the qShift and the mG + qShift models we used the empirical dataset to maximize the number of different time intervals and the number of occurrences within each interval so PyRate could properly estimate the preservation process. Because all occurrences have a temporal (i.e. dating) uncertainty, whose borders could potentially cross of the defined preservation time windows set, we built an R function (see below) that counts the number of occurrences inside each user-specified intervals and evaluated the balance and number of occurrences inside each of many possible set of intervals (supplementary figure 12). The desired is to maximize the number of occurrences completely

inside the defined intervals. As the possible combination of intervals are infinite, we took as candidates only conjunct of intervals classically associated with changes in faunal composition on the fossil record of mammal lineages. They were the GSA epochs (the early and late bounds of the paleocene, eocene, oligocene, miocene, pliocene and pleistocene), NALMA (North American Land Mammal Ages), ELMA (European Land Mammal Ages), SALMA (South American Land Mammal Ages) and ALMA (Asian Land Mammal Ages). As different datasets may have different intervals best fitted to data, we chose based only in the best interval for OUR dataset, which is the mainly reference dataset for all the comparisons in our questions. After the data analysis (results not shown), we determined that GSA epochs were the intervals which best fit to data, but then we tested some ages within each epoch to find a better fit to data. Our final choice was the following intervals (in mya): 56, 33.9, 23.03, 15.97, 11.65, 5.33, 2.58, 1.8, 0. They correspond, respectively to the Paleocene-Eocene boundary, the Eocene-Oligocene boundary, the Oligocene-Miocene boundary, the Burdigalian-Langhian boundary, the Serravalian-Tortolian boundary, the Miocene-Pliocene boundary, the Pliocene-Pleistocene boundary and the Gelasian-Calabrian boundary



Time (mya)

Supplementary figure 16. Best interval delimitation for a given dataset. Numbers in red show number of occurrences completely inside each interval, as calculated by our R function.

R function to determine the best fit of qShift intervals:

fossil_interv_finder=function(max, min, int){

```
#help:
#max = Occurrences late bounds (class: numeric)
```

```
#min = Occurrences early bounds (class: numeric)
#int = interval frontiers (class: numeric)
results=as.data.frame(matrix(data = NA, nrow = length(max), ncol = (length(int)-1)))
names=rep(NA, times=ncol(results))
#generating names:
for(i in 1:ncol(results)){names[i]= paste(int[i], int[i+1], sep="-")}
colnames(results)=names
#filling with
for(j in 1:(length(int)-1)){
    for(i in 1:length(max))
        if( min[i] >= int[j]-.2 & max[i] <= int[j+1]+.2) {results[i,j]=1}else {results[i,j]=0}
}
return(results)
</pre>
```

```
Usage example:
```

> counts = fossil_interv_finder(max=c(10,15,1), min=c(5,4,0), int=c(0,5,10)) #count which occurrences fall within each interval

> sum(apply(counts, 2, sum)) #This is the interval's score. The higher the better

Supplementary file 3

from "Ruminatia macroevolutionary patterns are robust to distinct taxonomic treatments on data from paleontological online databases"

The sampling of the Birth-Death reversible jump MCMC analysis:

We used the the birth-death model that uses a reversible jump Markov chain Monte Carlo algorithm (BDRJMCMC), recently implemented in PyRate, because it results in more accurate estimates than previous implementations (Silvestro *et al*, 2018). With the default parameters of PyRate we did not reached sufficient sampling (i.e. Effective sample size (ESS) > 200), even when using CAN dataset (results not shown), which were previously analysed with those older implementations (Cantalapiedra et al 2014).

We set the prior of the shape and scale parameters of the gamma-distributed prior on the preservation rate as 1.5 (both parameters), changed the fraction of birth-death rates updated at a time to 0.1, the frequency of updating shift times to 0.3, set the Frequency of model update in BDRJMCMC analysis to 10, set the Window size of updates of speciation/extinction rates to 2, and set the Update frequencies for preservation rate and birth-death parameters to 0.1 and 0.3, respectively. In PyRate notation this translates into the following example of python line:

 $\$ python PyRate.py OUR_sp_PyRate.py -N 197 -j 3 -out _0603 -mG -A4 -pP 1.5 1.5 -s 100000 -b 10000000 -n 110000000 -p 1000000 -fR 0.1 -fS 0.3 -M 10 -tR 2 -fU 0.1 0.3 0

You can check the meaning of the commands in the PyRate command list.

This sampling-associated changes (following Silvestro, personal communication) were chosen because in preliminary tests (results not shown) it improved considerably our effective sample sizes (ESS). Nearly all model parameters had ESS larger than 200 (dashed line) after this configuration (see figures below). We monitored ESS using the "ESS()" function in the package LaplacesDemon for the R environment, but some replicates were visually inspected using the software TRACER (Rambaut, 2015). The following figures (supplementary figures 17 - 36) show the ESS values for the most important PyRate parameters (parameter name is in y-axis). Note the following:

- 1. Given the setting of sampling frequency (put the value) and the length of the chain (put value) in our sampling strategy, 1000 is the maximum ESS value one analysis could have.
- 2. The PyRate parameter q in mG analysis is called "q_rate" and is estimated using the dataset in the full clade history. In qShift and mG+qShift analyses, q is estimated inside each predetermined interval, so there are multiples "qs", one for each interval. They are numbered "q_0", "q_1", ... untill the closest interval from the present.

- 3. The CAN dataset has only 8 windows of preservation because it did not have any occurrence in the earliest window, and therefore, in PyRate, it has values only from q_0 to q_7 . However, for visualization purposes, we paired the CAN parameters with the other datasets, so CAN dataset have q_1 to q_8 .
- 4. In qShift analyses, there is no "alpha" parameter given that this does not allow variation in preservation among species. In mG and mG+qShift analyses, "alpha" is estimated from the data.

References:

Rambaut, A., M. A. Suchard, D. Xie, and A. J. Drummond. 2015. Tracer v1. 6.

Silvestro, D., A. Antonelli, N. Salamin, and X. Meyer. 2018a: Improved estimation of macroevolutionary rates from fossil data using a Bayesian framework. BioRxiv 18:316992.

Supplementary figures (digital version)

Note: Because of printing limitations the ESS figures are available only in the digital version of the thesys. Supplementary figures 17 to 36 can be downloaded at: https://github.com/mjanuario/Taxonomy_Januario_Quental_sup_material/tree/master/sup_figs

Supplementary figure 17. ESS per replica for the q_8 parameter (mean preservation for occurrences before 56 Ma) in PyRate. Borders represents preservations models (brown for mG, black for mG+qShift and purple for qShift) and color of boxes represent the dataset (orange for ALL, blue for CAN green for OUR and red for RAW). The dashed line indicates ESS = 200.

Supplementary figure 18. ESS per replica for the q_7 parameter (mean preservation for occurrences between 56 and 33.9 Ma) in PyRate. Borders represents preservations models (brown for mG, black for mG+qShift and purple for qShift) and color of boxes represent the dataset (orange for ALL, blue for CAN green for OUR and red for RAW). The dashed line indicates ESS = 200.

Supplementary figure 19. ESS per replica for the q_6 parameter (mean preservation for occurrences between 33.9 and 23.03 Ma) in PyRate. Borders represents preservations models (brown for mG, black for mG+qShift and purple for qShift) and color of boxes represent the dataset (orange for ALL, blue for CAN green for OUR and red for RAW). The dashed line indicates ESS = 200.

Supplementary figure 20. ESS per replica for the q_5 parameter (mean preservation for occurrences between 23.03 and 15.97 Ma) in PyRate. Borders represents preservations models (brown for mG, black for mG+qShift and purple for qShift) and color of boxes represent the dataset (orange for ALL, blue for CAN green for OUR and red for RAW). The dashed line indicates ESS = 200.

Supplementary figure 21. ESS per replica for the q_4 parameter (mean preservation for occurrences between 15.97 and 11.65 Ma) in PyRate. Borders represents preservations models (brown for mG, black for mG+qShift and purple for qShift) and color of boxes represent the dataset (orange for ALL, blue for CAN green for OUR and red for RAW). The dashed line indicates ESS = 200.

Supplementary figure 22. ESS per replica for the q_3 parameter (mean preservation for occurrences between 11.65 and 5.33 Ma) in PyRate. Borders represents preservations models (brown for mG, black for mG+qShift and purple for qShift) and color of boxes represent the dataset (orange for ALL, blue for CAN green for OUR and red for RAW). The dashed line indicates ESS = 200.

Supplementary figure 23. ESS per replica for the q_2 parameter (mean preservation for occurrences between 5.33 and 2.58 Ma) in PyRate. Borders represents preservations models (brown for mG, black for mG+qShift and purple for qShift) and color of boxes represent the dataset (orange for ALL, blue for CAN green for OUR and red for RAW). The dashed line indicates ESS = 200.

Supplementary figure 24. ESS per replica for the q_l parameter (mean preservation for occurrences between 2.58 and 1.8 Ma) in PyRate. Borders represents preservations models (brown for mG, black for mG+qShift and purple for qShift) and color of boxes represent the dataset (orange for ALL, blue for CAN green for OUR and red for RAW). The dashed line indicates ESS = 200.

Supplementary figure 25. ESS per replica for the q_0 parameter (mean preservation for occurrences between 1.8 Ma and the present) in PyRate. Borders represents preservations models (brown for mG, black for mG+qShift and purple for qShift) and color of boxes represent the dataset (orange for ALL, blue for CAN green for OUR and red for RAW). The dashed line indicates ESS = 200.

Supplementary figure 26. ESS per replica for the *tot_length* (total branch length) parameter in PyRate. Borders represents preservations models (brown for mG, black for mG+qShift and purple for qShift) and color of boxes represent the dataset (orange for ALL, blue for CAN green for OUR and red for RAW). The dashed line indicates ESS = 200.

Supplementary figure 27. ESS per replica for the *root_age* (age of clade origin) parameter in PyRate. Borders represents preservations models (brown for mG, black for mG+qShift and purple for qShift) and color of boxes represent the dataset (orange for ALL, blue for CAN green for OUR and red for RAW). The dashed line indicates ESS = 200.

Supplementary figure 28. ESS per replica for the RJ_hp (reversible jump hyper-prior) parameter in PyRate. Borders represents preservations models (brown for mG, black for mG+qShift and purple for qShift) and color of boxes represent the dataset (orange for ALL, blue for CAN green for OUR and red for RAW). The dashed line indicates ESS = 200.

Supplementary figure 29. ESS per replica for the k_death (number of sampled extinction rate shifts) parameter in PyRate. Borders represents preservations models (brown for mG, black for mG+qShift and purple for qShift) and color of boxes represent the dataset (orange for ALL, blue for CAN green for OUR and red for RAW). The dashed line indicates ESS = 200.

Supplementary figure 30. ESS per replica for the k_{birth} (number of sampled speciation/origination rate shifts) parameter in PyRate. Borders represents preservations models (brown for mG, black for mG+qShift and purple for

qShift) and color of boxes represent the dataset (orange for ALL, blue for CAN green for OUR and red for RAW). The dashed line indicates ESS = 200.

Supplementary figure 31. ESS per replica for the *alpha* (shape parameter of the gamma distributed preservation heterogeneity) parameter in PyRate. Borders represents preservations models (brown for mG, black for mG+qShift and purple for qShift) and color of boxes represent the dataset (orange for ALL, blue for CAN green for OUR and red for RAW). The dashed line indicates ESS = 200.

Supplementary figure 32. ESS per replica for the q_rate (mean preservation rate for al time series) parameter in PyRate. Borders represents preservations models (brown for mG, black for mG+qShift and purple for qShift) and color of boxes represent the dataset (orange for ALL, blue for CAN green for OUR and red for RAW). The dashed line indicates ESS = 200.

Supplementary figure 33. ESS per replica for the *BD_lik* (likelihood of the birth-death) parameter in PyRate. Borders represents preservations models (brown for mG, black for mG+qShift and purple for qShift) and color of boxes represent the dataset (orange for ALL, blue for CAN green for OUR and red for RAW). The dashed line indicates ESS = 200.

Supplementary figure 34. ESS per replica for the PP_lik (likelihood of the preservation process) parameter in PyRate. Borders represents preservations models (brown for mG, black for mG+qShift and purple for qShift) and color of boxes represent the dataset (orange for ALL, blue for CAN green for OUR and red for RAW). The dashed line indicates ESS = 200.

Supplementary figure 35. ESS per replica for the *prior* parameter in PyRate. Borders represents preservations models (brown for mG, black for mG+qShift and purple for qShift) and color of boxes represent the dataset (orange for ALL, blue for CAN green for OUR and red for RAW). The dashed line indicates ESS = 200.

Supplementary figure 36. ESS per replica for the *posterior* parameter in PyRate. Borders represents preservations models (brown for mG, black for mG+qShift and purple for qShift) and color of boxes represent the dataset (orange for ALL, blue for CAN green for OUR and red for RAW). The dashed line indicates ESS = 200.

Supplementary tables

Database	Present in ALL	Present in RAW	PBDB	NOW
Markers of taxonomic uncertainty (occurrences removed)	Yes	No	"Cf", "cf.", "aff.", "?", "informal", "\"", "ex gr.", "? sp.", "sp. 1", "sp. 2", "indet.", "indeterminate", "sp.", "spp.", "reidentified", "form taxon"	"Genus id uncertain", "species id uncertain", "family id uncertain", "sp.", "indet", "indet.", "incertae sedis", "Findout", "taxonomic validity uncertain", "informal species", "genus attrib of species uncertain", "family attrib of genus uncertain"
Non Ruminant groups (removed)	No	No	Agriochoeridae, "Achaenodontidae", "Amphimerycidae", "Anoplotheriidae", "Anthracotheriidae", "Cainotheriidae", "Camelidae", "Cebochoeridae", "Choeropotamidae", "Diacodexeidae", "Dichobunidae", "Entelodontidae", "Dichobunidae", "Homacodontidae", "Leptochoeridae", "Merycoidodontidae", "Merycoidodontidae", "Mixtotheriidae", "Palaeochoeridae", "Protoceratidae", "Raoellidae", "Sanitheriidae", "Suidae", "Tayassuidae", "Xiphodontidae"	"Agriochoeridae", "Anoplotheriidae", "Anthracotheriidae", "Cainotheriidae", "Camelidae", "Dacrytheriidae", "Dichobunidae", "Entelodontidae", "Haplobunodontidae", "Hippopotamidae", "Merycoidodontidae", "Mixtotheriidae", "Palaeochoeridae", "Protoceratidae", "Raoellidae", "Suidae", "Tayassuidae", "Titanohyracidae", "Xiphodontidae"

Supplementary table 1: Markers in each dataset used to remove occurrences during data curation.

Supplementary table 2: Full list of synonyms used in OUR taxonomic treatment. This table has no printed version.

Supplementary table 3: Full list of comparisons between categories of names and datasets. This table has no printed version.

Supplementary table 4: Comparisons of taxonomical categories of all names within each dataset for species analysis. Note that not all categories are mutually exclusive (see supplementary table 3 for details on each name). This table has no printed version.

Supplementary tables 2, 3 and 4 can be downloaded at: https://github.com/mjanuario/Taxonomy_Januario_Quental_sup_material/tree/master/sup_tables

Capítulo 2

Re-evaluation of the "law of constant extinction" for ruminants at different taxonomical scales[†]

Matheus L. Januario¹ & Tiago B. Quental¹



1. Departamento de Ecologia, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil

[†] Manuscrito formatado de acordo com as normas da revista *Paleobiology*

Abstract

The "law of constant extinction", proposed by Van Valen, states that long and short-lived taxa have equal chances of going extinct. This pattern of age-independent extinction probability was inferred using the fossil record of several different taxa and has prevailed in evolutionary theory as one important assumption for decades. Its evidence relies on survivorship curves build from the literal reading of the fossil record, which fails to incorporate some important and inherent biases of the data. Additionally, Van Valen's analysis although mostly done at higher taxonomic level such as family and genus, has been used as a proxy to explain the dynamics at the species level. Using a very recent Bayesian framework, which accounts for several biases of the fossil record, including the fact that very short-living lineages might never make to the record itself, we showed that Ruminantia (Artiodactyla, Mammalia) species tend to present agedependent extinction, with a tendency for extinction probability to decrease with species age. We also analyzed the data at the genus level. Our results sugggest age-independence for genera although the extinction pattern seems less clear and potentially divergent from previous theory. We also found that the duration of a lineage is weakly related to the range size for species, while the duration of genera is related to its richness. Our results indicate that different taxonomic levels may present different extinction regimes, which could justify the develop of new macroevolutionary theory and methods.

Key-words: Age-independent Extinction, Extinction Biology, Macroevolution, Duration, Paleontology, Red Queen Hypothesis

Introduction

In 1973, Van Valen proposed the red queen hypothesis to explain a pattern that became to be known as "The Law of Constant Extinction". Using the fossil record of several lineages Van Valen (1973) proposed that within a given adaptive zone, the probability of extinction of a lineage does not depend on the lineage's longevity (i.e. for how long the lineage has existed). So at any given time, both short and long lived lineages within the same adaptive zone have equal probabilities of going extinct, all else being the same. This pattern was inferred using survivorship curves of higher taxa (mostly genera or families) within higher taxonomic groups (families, orders or phyla) by showing that those were essentially linear when plotted over a log scale for the ordinate axis. In this scale, a linear relationship mathematically describes extinction dynamics within a group as a decaying, memoryless process lacking an "aging" effect, constant within each particular higher group. Van Valen (1973) suggested that the slope of this linear relationship (and hence the average extinction rate) may vary across large groups, but that the linear relationship seemed to be an almost universal behavior for all biodiversity. Today, this age-independent extinction (hereafter only "age-independency") is a common assumption in many macroevolutionary models (e.g. Foote, 2003; Nee, 2006; Rabosky, 2006; Liow & Nichols, 2010; Morlon et al, 2011; FitzJohn, 2012; Silvestro et al, 2014a; Silvestro et al, 2014b; May et al, 2016; but see Ezard et al 2012; Alexander et al. 2016). The pervasiveness of this influence may be explained both by the consistency of Van Valen results and by its mathematical convenience for evolutionary methods.

It is important to note, however, that Van Valen derived his "law" of constant extinction mostly from empirical observations of constant probabilities for higher taxon (e.g. genus or families) within higher taxonomic ranks (e.g. families or Orders), but extended its implications to (and is used today as an assumption in) the species level dynamics within any adaptive zone. Although in his study he did look at species level in only four instances (planktonic foraminifera, dinoflagellates, and diatoms of the Centrales and Pennales groups), and taxonomic surrogacy is a common practice in paleontological studies (Raup & Sepkoski, 1986; Jablonski, 2017), this is in fact an open empirical question. Van Valen's choice of analyzing the data at higher taxonomic levels is a common paleontological practice justified by the fact that the fossil record is more complete at higher taxonomic levels (Sepkoski, 1992; Roy et al, 1996). It is also worth mentioning that at the time few methodological tools were available to explicitly incorporate the preservation biases inherent to the fossil record, which are specially problematic at the species level (Liow et al, 2011; Silvestro et al, 2014a; Silvestro et al, 2014b; Hagen et al, 2018). Moreover, methods employed by Van Valen (1973) were criticized (Raup 1975, Sepkoski 1975, Raup 1978), and recent re-evaluations of the pattern, at least for a small range of taxa where it has been done so far, found the opposite pattern: species present age-dependent extinction. While some of the recent results suggest age-increasing extinction probability (hereafter positive agedependency) for foraminifers (Pearson, 1995; Parker & Arnold, 1997; Doran et al, 2006; Ezard et al 2011; Ezard et al 2016), trilobites, conodonts and graptolites (Pearson, 1995), others suggest age-decreasing extinction probabilities (hereafter negative age-dependency) for carnivores (Hagen *et al*, 2018), Cenozoic North American mammals (Smits, 2015) and graptoloids (Crampton *et al*, 2016). To our knowledge, only one group tested at the species level showed evidence of age-independent extinction: calcareous nannoplankton (Pearson 1995). Apart from the criticism of Van Valen's methods and the contrasting evidence of later analyses at the species level, the Law of Constant Extinction continues to influence macroevolutionary theory as an almost universal assumption. Biological explanations for deviations from age-independency usually are related to an assumed increase (negative age-dependency) or decrease (positive age-dependency) in "fitness" or range size.

There are practical reasons why there are few macroevolutionary studies testing the law of constant extinction in the fossil record at the species level, all related to the fact that the fossil record is incomplete. First, the methods used to answer this question need a large amount of data (i.e. longevities of many species), and the fossil record of many groups does not meet this demand. Second, most previous work that addressed this question had taken a literal reading of the fossil record (i.e. the use of the first and last occurrence data to define species longevities). This assumes that imperfections of the fossil record are not relevant, which is unlikely to be true for most clades. The literal reading of the fossil record becomes even more problematic when a considerable number of species have only one occurrence (i.e. singletons). Hence taking a literal reading of the fossi record will strongly distort our estimates of taxon longevity, except for lineages which the record is extremely good (e.g. Foraminifera). In fact true longevities are very difficult to estimate because the first and last occurrences are unlikely to represent the true times of origin and extinction (Liow et al, 2011; Silvestro et al, 2014a; Silvestro et al, 2014b). The literal reading of the fossil record then truncates the longevities at different levels according to the record quality and fossil preservation biases. Moreover, it is also likely that species that have lived for very short time interval do not even make it to fossil record, which result in a biased sample of species with medium and long longevities. Although analyses of higher taxa ameliorate this sampling problem, the idea that they could be used as proxies for species level dynamics (an assumption made by Van Valen, 1973), is controversial among paleontologists, especially when the number of species is much larger than the number of higher taxa (Roy et al, 1996).

Theoretically, there is even less reason to think that the extinction regime at one taxonomic level matches the extinction regime at other levels of biodiversity. This is because simple birth-death models predict that if the species dynamics shows age-independency in extinction probability, the extinction probability of the higher taxa tends to decrease with time since their origin (Raup 1978). This pattern holds true for any combinations of death (extinction) and birth (speciation/origination) rates, but the pattern is accentuated for lineages with greater positive diversification rates, when birth rates are significantly higher than death rates, because here "sub lineages" quickly accumulate, considerably lowering the extinction probability of their higher group (Raup 1978). In fact, previous empirical work (Anstey, 1978; Raup, 1978; Finnegan *et al*, 2008) has systematically found age-dependency with age-decreasing rates of

extinction in the genus level, while the results at the species level seem more variable (at least in respect to the direction of the age-dependency) as mention earlier. Given Raup's (1978) predictions of different patterns of age-dependency at different taxonomic levels, surprisingly few studies have tested age-dependency for the same clade at different taxonomic scales. To our knowledge, the only attempt to try such comparison was a simple illustrative example conducted by Ezard *et al* (2016), which found that different taxonomic scales (species vs genera) of macroperforate planktonic foraminifera may have divergent patterns of age-dependency. The present work has two goals. The first is to test the hypothesis of age-independency at the species level. The second is to investigate if the pattern of age-dependency hypothesis at both the species and the genera taxonomic levels for all Ruminantia (Mammalia), a monophyletic mammalian lineage with a fossil record that is good enough for such analyses (Januario and Quental in prep., chapter 1).

Methods:

We used the downloaded and curated dataset described by Januario & Quental (in prep., chapter 1 - details in the Supplementary file 1), at the species level and the same dataset collapsed to the genus level (for details in data curation, dataset generation and taxonomy properties, please see Januario & Quental, in prep., chapter 1).

Fossil occurrences are not precise estimates of spatial-temporal data, but rather represent a time interval associated to the strata where the fossil was found. To account for the age uncertainty in each fossil occurrence 100 replicates were generated by randomly sampling a point time estimate for each occurrence temporal range. Each one of those 100 replicates was then used in the sub-sequent diversification analysis. To account for the fact that the observed times of origin and extinction of a given lineage on the fossil record likely represent an underestimate of the true times of those events, we analyzed the data using a birth-death reversible jump markov chain monte carlo (hereafter: birth-death estimation) which simultaneously estimates the duration of all lineages sampled in the dataset, the birth-death parameters and the preservation processes. More specifically, we followed the framework implemented in Hagen *et al*, (2018) to infer the signature of age dependence (or independence) over extinction in ruminants. This approach is implemented in the PyRate software (Silvestro et al, 2014a; Silvestro et al, 2014b; Silvestro et al, 2018; Hagen et al 2018). In this framework, fossil occurrences are used to simultaneously estimate (1) the "true" (i.e. uncertainty-informed) times of origination/speciation and extinction of all lineages sampled within the fossil record, (2) the rate of preservation (which includes fossilization, sampling and identification in a single parameter) within the fossil record and (3) speciation and extinction rates through time. This framework uses all occurrences available (in contrasts with other methods that generally only use the first and last occurrence of each lineage) to estimate the diversification parameters, which can improve parameter estimation and at the same time calculate the uncertainty of the estimates. More importantly, the method models the longevities of species that did not last long enough to

be present in the fossil record. This is relevant given that not taking those into account might change the distribution of species longevities and hence the inference of age-dependency or age-independency.

Using the PyRate framework, we fitted an age-dependent extinction model (Hagen *et al*, 2015; Hagen *et al*, 2018). Under this model, the time until extinction (the species longevities) of all lineages in the dataset follows a Weibull distribution where the shape parameter describes the effect of the age dependence. For a shape parameter <1, the extinction is higher for young lineages, while for shape parameter >1 extinction probability increases with taxon age. If shape parameter approximate 1, then extinction is age-independent, a pattern congruent to what was found by Van Valen (1973) (supplementary figure 1). In this last case, the Weibull distribution reduces to an exponential distribution, which characterizes a simple birth-death model, commonly assumed in the macroevolutionary practice in general.

Due to the nature of the fossil record, preservation potential might vary substantially among different geological times. This bias can strongly affect the shape parameter estimations, so our method explicitly accounts for preservation heterogeneity between different geological periods in time. To account for such temporal variation in preservation, we allowed preservation to vary between geological periods but remain constant within each period. As the duration of lineages is directly determined by the extinction rate (Marshall, 2017), any shifts in the extinction rate will combine species under very different extinction regimes and hence very different distributions of lineages duration. To avoid the influence of this effect, we restricted our analyses to intervals of time when the extinction rate was constant (Supplementary figure 2). Following the diversification dynamics found by Januario & Ouental (in prep.; chapter 1), and some data properties (see chapter 1), we used the time since the root age of the Ruminants (estimated from data using PyRate) until 10.5 mya as the window for the genera-level analysis, a period with one extinction regime. For the species-level analysis, we used the interval between 39.5 and 8 mya as a conservative time window for the analysis. For the species level analysis we also estimated the parameters for a wider window between the Ruminantia root age (estimated from data using PyRate to be around 58 Ma) and 6.5 mya. The difference between the wider window and the conservative analysis is essentially driven by the significance of rate shifts (Januario and Quental in prep., chapter 1). The conservative window was determined by limiting the time window to a period that does not include any evidence of extinction rate shift, while the wider window includes extinction rate shifts with lower significance. None of them include extinction rate shifts that were highly significant (Bayes factor > 6; for details see Supplementary file 1). We have a total of 82 lineages and 767 occurrences in the genera analysis, 287 lineages and 1244 occurrences in conservative species analysis, and 361 lineages and 1704 occurrences in wider window species analysis. In the PyRate framework, each fossil occurrence has a number of replicates which were generated by randomly sampling a point time estimate (inside the timespan between early and late bounds) for each occurrence. PyRate estimated the duration only for lineages which had all their point time estimates of that replica inside the pre-defined time window used, otherwise we would inadvertently truncate some longevities. So the composition

of species in which the durations were estimated may vary from one replica to another, and not all (although most) species will be included in all replicas of a given dataset.

To guarantee sufficient sampling (ESS > 200) in most replicas, we ran the RJBDMCMC (hereafter birth-death) integrator using the exact options used in Januario and Quental (in prep.; chapter 1), with only the following differences: i) we run a total of 50 000 000 iterations, ii) discarded the first 1000 iterations as burn-in, iii) sampled every 25 000 iterations to obtain the posteriors for each parameter and each replica, iv) set the age-dependency model described above to within the time window described above. We then compiled the results for all the replicas within a single posterior distribution to summarize the parameter estimation (shape and scale) of the Weibull-modeled distribution of lineage durations in the Ruminantia fossil record. Because of size constraints in log files, after the burn-in step, we resampled (i.e. pruned) 200 of 1000 sampled iterations of each replica to create our final dataset for all analyses. We inferred if the shape parameter of the ruminant duration distribution was significantly different from 1 if its 95% highest posterior density (HPD) interval did not overlap with 1.

To access the role of range size on modulating the duration of a lineage, we made a simple linear regression of estimated lineage duration (we used the median of the genera duration posterior distribution) as a function of lineage range area. To calculate range area, we concatenated all temporal occurrences of a given sampled lineage (for all genera and species datasets) and used the function CalcRangeSize() from the package speciesgeocodeR (Töpel, *et al*, 2016). We discarded ranges of species with less than 3 occurrences. Later we log-transformed (base 2) the estimated area. This is a simplistic (but still efficient in relation to other methods – Darroch & Saupe, 2018) way to estimate range size because its pattern is known to vary through time within lineages (Foote *et* al, 2008; Žliobaitė *et al*, 2017) and also because the concatenation of temporal occurrences is an extremely simple method of range calculation which does not account for the inherent biases (as said above) of the fossil record. Yet another limitation here is that we do not correct for phylogenetic relationships, given that no species level phylogeny that includes all extinct Ruminantia exists.

According to Raup (1978) age-dependency at higher taxa could be explained not only by the extinction (i.e. duration) pattern of its constituent lineages, but also because of their pattern of speciation/origination, as a higher taxa has its extinction time only with the extinction of its last species, which means there is a role for speciation on determining the duration of genera. This explanation leads to a simple testable prediction: richness should be a good explanatory variable of higher taxa duration. Even though the birth-death estimation implemented here takes into account the unsampled lineages when modeling the age-dependency (weibull's shape and scale) parameters, it does not explicitly models diversity through time accounting for the unsampled lineages. Hence this association is limited to the sampled species assuming that at least the relative rank of species richness is captured by the sampled species. We tested this prediction using a simple linear regression in which sampled richness predicts genus duration (we used the median of the genera duration posterior distribution). In this regression, the intercept was forced to zero as a genus with zero species has no duration.

Results:



Figure 1. Violin density plot of the posterior distribution estimated for the shape parameter in all three datasets. The dashed line represent a shape parameter equals to 1.

The ESS associated with the birth-death run for each replica indicated that we reached sufficient sampling (ESS > 200) in all datasets and windows (see supplementary figure 8, 9 and 10). Our analysis (figure 1) found strong evidence for age-dependence in species level, as both the conservative (median = 0.66, 99% HPD = 0.48 - 0.87) and the wider window analysis (median = 0.66, 99% HPD = 0.49 - 0.84) had their estimated shape parameter of the Weibull distribution significantly smaller than 1. The estimated scale parameter (conservative analysis: median = 2.06, 95% HPD = 1.26 - 2.83; wider window analysis: median = 2.11, 95% HPD = 1.38 - 2.82) and the estimated species mean longevity (conservative analysis: median = 2.74, 95% HPD = 2.19 - 3.32; wider window analysis: median = 2.84, 95% HPD = 2.31 - 3.35) were also similar for both datasets. The coefficients of variation of the means of shape (Conservative = 0.057; Wider window = 0.051) and scale (Conservative = 0.051; Wider window = 0.085) parameters estimated across the different replicas were very small. This, together with our visual inspection of the shape estimates for each replica (see supplementary figure 11 for conservative analysis, supplementary figure 12 for wider window analysis) indicates that the overall shape estimative is made of a consistent pattern within each replicate rather than differences between replicates, which indicates that the specific configuration of the sampled replicas accounted for little variation in the estimates. A visual inspection of the overall shape estimative suggests that the Weibull distribution estimated including non-sampled species seems well adjusted to the

distribution of empirical sampled species also estimated from PyRate (figure 2). It also shows that species with very short longevities (i.e. durations smaller than 1 my) were particularly absent from the empirical dataset (note the small density at the left of the distribution on figure 2).

In the genus level analysis, the results indicated the opposite pattern; we found strong evidence for AIE, both the 95% HPD interval estimated for the shape parameter (95% HPD 0.73 - 1.27) crosses 1 (figure 1), and the median of the posterior distribution of the shape parameter (= (0.99) is very close to 1. The scale parameter in the genera level (median = 6.14; 95% HPD = 4.47 - 7.82) is also estimated as substantially larger than in the species datasets, as the mean genera duration estimated for the dataset (median = 6.15; 95% HPD = 4.46 - 7.82). The coefficients of variation of the mean shape (CV = 0.057) and scale (CV = 0.052) parameters estimated across all replicates were small. This, together with the similarity between the shape posteriors (see suplementary figure 13) indicates that the large HPD interval is possible due to uncertainty within each replicate rather than differences between replicates. Although the median of the posterior distribution of the shape parameter is very close to 1, the large HPD interval of the genus analysis is almost 50% larger than in the species level. The variation of estimates can also be visualized by plotting the expected curve given the mean estimated shapes of the Weibull distribution over the distribution of genus empirical longevities, also estimated from PyRate (figure 3). There are two worth mentioning deviations of the empirical distribution from the expected distribution for a weibull with shape parameter equal to 1: one is a large absence of durations between 1.2 and 4 My, and the second is an accumulation of durations between 5 and 6.8 My. Those characteristics make the distribution to be far from a typical weibull curve (figure 3), which might explain the large uncertainty in shape estimate for the genus level. It is also worth mentioning that when we look at the nature of the genera, those monotypic genera (light grey on figure 3), show as expected a distribution more akin to the one found for species (see figure 2), while the genera with multiple species seems to have a distribution that would likely have a shape parameter higher than 1.



Figure 2: Empirical species duration distribution estimated in PyRate for the sampled taxa. Dark grey bars indicate the species which at least one occurrence was outside our time window analysis in at least one of the PyRate replicas, and light grey indicate species with all occurrences completely inside the window. To estimate the shape and scale parameters, only species with all occurrences fully inside the time window, for that given replicate, were used. Panel A shows the conservative analysis and panel B show the wider window analysis. The Weibull distributions estimated for different iterations by PyRate including the estimation of non-sampled lineages are represented by red lines (600 randomly drawn iterations from the combined posterior which had a total of 20000 iterations = 200 for each of the 100 replicates. The number 600 is the maximum number allowed by ggplot2).



Genus class Politypic Monotypic

Figure 3: Empirical genus duration distribution estimated in PyRate for the sampled taxa, including the genera which were outside our time window analysis in at least one of the PyRate replicas, and the Weibull distributions estimated for different iterations by PyRate including the estimation of non-sampled lineages (red lines – 600 randomly drawn iterations from the combined posterior which had a total of 20000 iterations = 200 for each of the 100 replicates. The number 600 is the maximum number allowed by ggplot2). The dark bars indicate genera with more than one species, and the light bar indicate monotypic genera.

Our linear regressions show that range area is related (although marginally in the case of wider window) to lineage duration in species level, but it does not explain much of the duration variation (conservative: figure 4.A, R^2 = 0.11, Intercept = 2.15, slope = 0.23, t = 3.41, p = 0.0009; wider window: figure 4.B R²= 0.05, Intercept = 3.53, slope = 0.16, t = 2.26, p = 0.0257). The linear models also show that range size is not related to duration in the genus level (Figure 5.A, R²= 0.04, Intercept = 4.46, slope = 0.30, t = 1.50, p = 0.13) levels. The linear regressions also show that there is an association between genera richness and duration (figure 5.B; R² = 0.31, slope = 2.37, t = 12.85, p = <2e-16).
Species level - Conservative



Figure 4: Linear model in which the median of the posterior duration distribution is predicted by the logtransformed range size for the sampled Ruminantia species in the conservative (panel A) and wider window (panel B) analyses. The red dots represent species which are present within the respective analysis window in all PyRate replicas. The black dots indicate the species which were outside the respective time window analysis in at least one of the PyRate replicas (only for showing purposes). The lines represent the linear model estimated using only the genera present in all replicas and the 95% confidence interval is represented by the hatched area.



Figure 5: Linear model in which the median of the posterior duration distribution is predicted by the logtransformed range size (panel A) the genera richness (panel B – jittered in the abscissa) for the sampled Ruminantia genera. The red dots represent genera which are present within the analysis window in all PyRate replicas. The black dots indicate the genera which were outside our time window analysis in at least one of the PyRate replicas (only for showing purposes). The lines represent the linear model estimated using only the genera present in all replicas and the 95% confidence interval is represented by the hatched area.

Discussion:

By explicitly taking into account the imperfection of the fossil record we have shown that extinction probability of Ruminat species decreases as species live longer, contradicting Van Valen (1973) expectation of age-independent extinction. Although age-independent extinction within an adaptive zone has been termed a pervasive evolutionary law by Van Valen, there are surprisingly few studies that have in fact looked at this question at the species level (including Van Valen's work for most part). This absence is tightly linked to the biases one is subject by analyzing species longevities at face value. Not only first and last occurrences very likely represent poor estimates of the true times of origination and extinction, but short-lived species are very likely to not even make to the fossil record. The few exceptions are those invertebrate lineages (e.g. Forams and graptoloids) with a relatively complete fossil record (e.g. Pearson, 1995; Doran *et al*, 2006 Ezard *et al*, 2012, Crampton *et al*, 2016). More recently, the development of new methods that explicitly take into account the imperfection of the fossil record (the method used here), has allowed taxa with poorer records to be analyzed at the species level, and so far this has been done only with mammals (Hagen *et al*, 2018; and this study).

Contrary to Van Valen's expectation, most studies at the species level have found agedependent extinction, either positive or negative. Examples of negative age-dependency are: species of ruminants (this study), graptoloids (Crampton *et al*, 2016), Cenozoic North American mammals (Smits, 2015) and carnivores (Hagen *et al*, 2018). The opposite pattern - that is, extinctions probabilities that increase with age in species - has also been found for other groups (or in the same group in different moments in time - Crampton *et al*, 2016). Those include planktonic foraminifera (Pearson, 1995; Parker & Arnold, 1997; Doran *et al*, 2006; Ezard et al 2011, 2016), conodonts, graptolites and trilobites (Pearson, 1995). The few examples of ageindependency at the species level are: planktonic foraminifera, dinoflagellates, and diatoms of the Centrales and Pennales groups, those all investigated by Van Valen (1973), with Pearson (1995) founding the only age-independent example wich is not from Van Valen's investigation: calcareous nannoplankton.

There are at least three explanations that have been proposed for why species would present negative age-dependency. First, it is assumed that as species age, their geographical ranges increase, and then larger ranges result in lower extinction risk (Payne & Finnegan, 2007; Foote et al 2008; Jablonski, 2008). This has been shown for some invertebrate and few vertebrate lineages (Foote et al 2008; Jablonski, 2008). Our results show a weak correlation or no correlation at all between range size and duration for species (figure 4) and genera (figure 5) (but note that our analysis represents only a simplistic view on the range dynamics through time). This result could also be influenced by the fact that species which are known from a limited number of occurrences may bias the estimate of ranges towards small values. A better understanding of this pattern would benefit from taking in consideration not only temporal but spatial incompletness of the fossil record, and the possibility of temporal variation in range size.

This is a methodological challenge (but see Žliobaitė *et al*, 2017), but a major concern if one is interested in documenting how a species occupies different localities at different points in time.

Secondly, similar to Raup's argument of age dependency for the higher taxa, a hierarchical effect could also play a role at the species level. Just as the probability of survival increases as a clade last longer by the generation of new species, a species might last longer due to the multiplication of its subunits (e.g. populations or metapopulations). This possibility would mean that the birth-death process is a multiplicative process (i.e. every birth gives rise to a new birth-death process) and uses the same logic of Raup's model but in a lower level processes within species: the birth and death of populations. This is theoretically possible as the mathematical properties are the same in multiplicative and non-multiplicative birth-death process (as shown by Bailey, 1964). This possibility also allows the birth and death of populations to be age-independent, but this is an open empirical question. Despite this possibility, finding paleontological evidence for this pattern is challenging due to the incompleteness of the fossil record, which makes it difficult to describe the temporal dynamics of populations over geologic timescales. It is also challenging to disentangle this hierarchical effect from the range effect mentioned earlier given that the addition of new, disconnected, populations also results in range expansion.

A third possibility would be the evolution of extinction resistant traits which then cause the age-dependency. It is important to note that, as proposed by Finnegan et al (2008), the decrease in extinction rate as lineages get older could result from an overall increase in the fitness within each lineage (i.e. anagenetically) or through the higher-level selection ("culling" sensu Finnegan et al, 2008) of the "less fit" lineages. Although interesting, these underlying mechanisms of trait evolution are challenging to tease apart without detailed phylogenetic information for the clade studied and a candidate trait mechanistically liked to an increase in species fitness (e.g. the traits cited in Jablonski, 2008) at the species level. The work done by Boyajian & Lutz (1992), which used complexity in ammonite shell sutures as a proxy for biological complexity, where the authors argue for complexity being potentially a extinction resistant trait, suggested that an increase in shell suture complexity was not related to extinction probability. On the other hand, Smits (2015) showed that more generalist species of Cenozoic North American mammals had a lower extinction probability suggesting that indeed a extinctionresistant trait might evolve. Furthermore, the results of Crampton et al (2016) indicate that agedependency could (and in the graptoloid case do) vary through absolute geological time, passing through moments of positive and negative age-dependency, and even through moments of ageindependency. If common for other taxa, this time variation in age-dependency regime could lead to future relevant insights over general rules regarding extinction regimes, in particular the prevalent direction of age-dependency.

Altought species are the canonical unity of interest in macroevolution, most of the paleontological investigations that tested for age-independency over extinction using multiple clades did it at the level of higher taxa (e.g. Van Valen, 1973; Boyajian, 1986; Finnegan *et al*, 2008). This decision was very likely influenced by the inherent biases of the fossil record, and

the belief that higher taxonomic levels present much better preservation. Apart from the empirical question if higher taxa is indeed a good proxy for species level dynamics, one should note that many of the generalizations or predictions made about species dynamics relied on analysis made with higher taxa (Van Valen's work being the most prominent of those). Raup (1978) provides a mathematical model which indicates that genera and species may have inherent differences in their age-dependency pattern, at least in the case species present age-independency. The framework implemented here tries to account for the preservation biases within the fossil record, making it possible to simultaneously investigate extinction age-dependency at the species level and genus level, and hence making it possible to explicitly compare the patterns at both taxomonic scales for the same clade.

The Ruminantia extinction pattern found at the genus level, although statistically not different from an age-independent regime, is uncertain as the HPD interval of the Weibull's shape parameter is very broad and includes values consistent with positive and negative agedependency. Is important to note that these results may be affected by the violations of different assumptions of the method, but the two most obvious assumption violations were probably taken into account, at least partially. Those important assumptions are: (1) no time variation in extinction rate and (2) abscence of considerable pseudo-extinction. Those were taken in to account, at least partially, by the use of time windows on which extinction rate was constant (assumption 1), and by the taxonomical review made on data (Januario & Quental, , chapter 1) prior to analysis (assumption 2; but note that only a phylogenetyc analysis would properly mitigate the effect of pseudo-extinction). We also interpret as unlikely that the genera-level large HPD intervals are hardly influenced by the incompleteness of the fossil record alone, as the mean number of occurrences per lineage are two times higher for the genus level than for the species level analysis. If the incompleteness of the fossil record alone influenced our results, we would expected that the precision (i.e. the length of the HPD interval) would be related to the number of occurrences per lineage, and we found the opposite: the species-level analysis have shorter HPD intervals and less occurrences per lineage. This conclusion is also reinforced by the lack of correlation between preservation rate (q parameter in PyRate) and the shape parameter (Supplementary figures 14, 15, 16, 17 and 18).

Ezard *et al* (2016) found a similar broad estimate for the parameter which describes the age-dependency behavior at the genus level and interpreted the broad range as a result of the small sample size (smaller number of logevities used in the genus level analysis). Altought we also did not have a very large number of logevities estimates at the genus level , we suspect that other factors might also have played a role here. In particular, we highlight that the distribution of Ruminantia empirical genus longevities seem to have a bi-modal distribution (figure 3). Additionally, if we visually compare (no proper test was done) the monotypic genera to genera with multiple species, the former seems to better conform to a negative age-dependent extinction, while the later to a positive age-dependency (figure 3). Hence it is possible that this multi-behavior generates the uncertainty in the shape parameter estimate at the genus level. Hence our genus level result suggests that some lineages might not conform to simple regime of

either positive or negative age-dependency, but rather to a combined effect of both. If this is the case another distribution rather than the Weibull should be used in further researches. This would also imply that different mechanisms may pace extinctions at young (e.g. species more prone to demographic effects associated with small population or range size) or old (e.g. a ratched mechanism that diminishes evolvability) ages. It is also possible that this bi-modal distribution of longevities at the genus is produced by shifts in the species dynamics. Even thought we choose a time window where neither origination nor extinction changed through time at the genus level, at the species level this time window included a shift of extinction rate (see chapter 1). Hence it is possible that our colletion of longevities at the genus level includes two distinct species-level regimes. Given that the different longevities at the genus level is a product of the the dynamics at the species level, it is possible that the bi-modal distribution at the genus level is the result of those two distinct diversification regimes at the species-level. If this represents a strong bias, it also could affect other papers that found age dependency at the genus level (e.g. Jones & Nicol, 1986; Boyajian et al, 1992; Finnegan et al, 2008) because those did not analysed the dynamics at the speciel level and could therefore be incurring in the same problem. One should also note that when analysing longevities at the genus level, not only changes in the extinction rate at the species level should be taken into account, but also potential changes in speciation rate at the species level.

Irrespective of what drives the uncertainty at the Ruminantia genus level, other investigators have found a consistent pattern for negative age-dependency on higher taxa. The extinction probability tends to decrease with age in higher taxa of clades such as bryozoans (Anstey, 1978), ammonoids (Boyajian *et al*, 1992), rudists (Jones & Nicol, 1986), and several marine organisms (Raup, 1978; Boyajian, 1986; Boyajian 1991; Boyajian 1992; Finnegan *et al*, 2008). Hence evidence for age-independency seems fragile also at the genera level, as it has been only found by (1) Van Valen (1973) without an access to uncertainty, (2) by Ezard et al (2016) and (3) by our results for Ruminantia, the last two presenting very broad uncertainties (much broader than at the species level).

According to Raup (1978) age-dependency at higher taxa could be explained not only by the extinction pattern of its constituent lineages, but also on their pattern of speciation/origination. Our results from the linear model indicate that Raup's mechanism may have an effect on the Ruminantia genera dynamics, as genera duration is related to genera richness (figure 5B). Hence, to some extent this result partially supports Raup's prediction where genus longevity is influenced by species dynamics, even though we did not find any strong support for age-dependency at the genus level. We also found, with our simplistic analysis, a very weak association between longevity and range size at the species level and no association at the genus level. It is interesting to note that the effect of range size in lineage duration, if present in species, is very different between the two taxonomic levels, which could indicate that the effect of range on extinction is dependent on the taxonomic scale an investigator analyses, at least in the ruminant case. Another possibility is that range is not related to duration in both taxonomic levels and our analysis was influenced by some bias, such as the propensity of lineages with fewer occurrences to also be more spatialy restricted. In this scenario the weak association between species duration and geographical range would be produced by a third variable that affect simulatenously duration and geographical range, which is variation in preservation among species.

The predictions of Raup's model that at higher level (e.g. genus) we should find agedependency if lower level (e.g. species) presents age-independency, and the discrepancy between the genera and species age-dependency pattern for higher taxa in our results and in Ezard et al (2016) raise questions over the correspondence of species, genera and other higher taxa in macroevolutionary dynamics. Although the pattern of a higher level is historically assumed to be a sufficient proxy for lower levels (Van Valen, 1973; Raup & Sepokoski, 1986; Lu et al, 2006; Jablonski, 2017), different taxonomic levels seem to capture different aspects of biological diversification (Ezard et al, 2016). Altought Raup's (1978) model provides testable predictions for why different taxonomic scales have different durations (e.g. why higher taxa might be agedependent when sub-units are age-independent), the few empirical evidence already available suggests that the process described by Raup (1978) may be relevant but insufficient to explain higher taxa duration distribution, for example in our analysis species level shows agedependency and genus level shows either a age independency or perhaps a mixed behavior of positive and negative age-dependency. Also, Raup (1978) did not formalized what are the predictions for higher taxa (e.g. genera) when subunits (e.g. species) have age-dependent extinction, and this is a possible future investigation to be taken.

Taken our result together with those of other studies we suggest that, in ruminants but possibly in other biological groups, age-dependency seems to be the rule at species level and that the genus level analysis might not always reflect directly the species level dynamics. But to what extend this results affect the status of the Red Queen hypothesys, appointed by Van Valen (1973) as the explanation for a pervasive, multi-scale age-independency across all biodiversity? The very recent review by Strotz et al (2018) argues that the absence of age-independent extinction inpugns Van Valen's Red Queen Hypothesys as a valid explanation for macroevolutionary patterns observed throught the fossil record, because it woud only explain the pattern described by the Law of Constant Extinction. So before "dethroning the Red Queen out of its kingdom", one first needs to answer if the underlying mechanism of the Red Queen Hypothesys could also make sense in explaining a pattern of age-dependent extinction. For example, negative agedependency does not reject directly the "zero-sum-like" nature of biological interactions shaping macroevolution, as it is possible that few long-lived lineages maintain, during the majority of time, resources to the detriment of other lineages (generating, for example, incumbency), and so the dissapearance of a lineage may influence the extinction probability of other remaining lineage (Sepkoski et al, 2000; Liow et al, 2015; Silvestro et al, 2015), which is the expected under a zero-sum game (we highlight that the zero-sum game do not need to have players with equal probabily to "win" the game). The other component of the Red Queen rationale, the necessity to "keep pace" with a constantly changing environment (which is made of abiotic and biotic factors such as biological interactions) is still qualitatively predicted by the single fact that no lineage lasts forever, a fatality of any regime of extinction regardless of its age-dependency or independency. The only difference between the scenario proposed here and Van Valen's proposition is that in this case a zero-sum game associated to the neccessity to "keep up with the pace" of environmental change would also lead to some kind of progress with respect to the lineage's ability to respond to extinction risk. Curiosly, for marine invertebrates, Alroy (2008) suggested that the secular trend of extinction since the Cambrian is one of decrease in extinction rate.

Although age-dependency might be the most comon extinction regime at species level, the direction of the age-dependency seems quite variable. Different biological groups present different directions of age-dependency, different taxonomic levels might differ with respect to the age-dependency pattern (Ezard et al 2016; this study), and there is even temporal variation within groups for the prevalence and direction of age-dependency (Crampton *et al*, 2016). One could argue that this variation is related to the different approaches used in the literature to evaluate age-dependency or its absence, but the group most used in such analyses, planktonic foraminifera, has found to present a consistent pattern of positive age-dependency, even under different approaches, data, and even species concepts (Pearson, 1995; Parker & Arnold, 1997; Doran *et al*, 2006; Ezard *et al*, 2016). Hence we suspect that this variation is real and it calls for the re-consideration of age-independency as a general assumption in macroevolution and justifies the development of more flexible birth-death models in the field (e.g. Goldberg *et al* biorXive).

Acknowledgements: We thank Daniele Silvestro for help with the interpretation of PyRate parameters and Paulo Guimarães Jr. and Eduardo Santos for the provided suggestions at different stages of this study. M.L.J. was funded by Conselho Nacional de Desenvolvimento Científico (CNPq), Brasil (process number 130126/2017-9.). T.B.Q. would like to thank FAPESP for financial support (grants # 2012/04072-3 and #2018/05462-6)

Competing interests: The authors declare no competing interests

Code availability: The scripts used can be downloaded at the authors github https://github.com/mjanuario/ADE_Januario_Quental_sup_material/tree/master/scripts

References:

- Alexander, H., A. Lambert, A. Modeles, and T. Stadler. 2014: Quantifying age-dependent extinction from species phylogenies. Systematic biology. 65:35-50
- Alroy, J. 2008. Dynamics of origination and extinction in the marine fossil record. Proceedings of the National Academy of Sciences 105:11536-11542

Anstey, R. L. 1978: Taxonomic survivorship and morphologic complexity in Paleozoic bryozoan genera.

Paleobiology 4:407–418.

- Bailey, N. T. J. 1964. The Elements of Stochastic Processes. 249 pp. John Wiley & Sons, Inc..: New York, NY.
- Benton, M. J. 2016: Origins of biodiversity. Palaeontology 56:1-7.
- Boyajian, G., and T. Lutz. 1992: Evolution of biological complexity and its relation to taxonomic longevity in the Ammonoidea. Geology 20:983–986.
- Boyajian, G. E. 1992: Taxon age, origination, and extinction through geologic time. Historical Biology 6:281–291.
- Boyajian, G. F. 1991: Taxon Age and Selectivity of Extinction. Paleontological Society 17:49-57.
- Boyajian, G. F. 1986: Phanerozoic trends in background extinction: consequence of an aging fauna. Geology 14:955–958.
- Brockhurst, M. A., T. Chapman, K. C. King, J. E. Mank, S. Paterson, and G. D. Hurst. 2014: Running with the Red Queen: the role of biotic conflicts in evolution. Proceedings of the Royal Society B: Biological Sciences 282.
- Crampton, J. S., R. A. Cooper, P. M. Sadler, and M. Foote. 2016: Greenhouse-icehouse transition in the Late Ordovician marks a step change in extinction regime in the marine plankton. Proceedings of the National Academy of Sciences 113:1498–1503.
- Darroch, S. A. F., and E. E. Saupe. 2018: Reconstructing geographic range-size dynamics from fossil data. Paleobiology 44:25–39.
- Doran, N. A., A. J. Arnold, W. C. Parker, and F. W. Huffer. 2006: Is extinction age dependent? Palaios 21:571–579.
- Ezard, T. H. G., T. B. Quental, and M. J. Benton. 2016: The challenges to inferring the regulators of biodiversity in deep time. Philosophical Transactions of the Royal Society B: Biological Sciences 371:20150216.
- Ezard, T. H. G., P. N. Pearson, T. Aze, and A. Purvis. 2011: The meaning of birth and death (in macroevolutionary birth death models) Subject collections The meaning of birth and death (in macroevolutionary birth death models). Palaeontology:139–142.
- Finnegan, S., J. L. Payne, and S. C. Wang. 2008: The Red Queen revisited: reevaluating the age selectivity of Phanerozoic marine genus extinctions. Paleobiology 34:318–341.

Fitzjohn, R. G. 2012: Diversitree: Comparative phylogenetic analyses of diversification in R. Methods in

Ecology and Evolution 3:1084–1092.

- Foote, M. 2003: Origination and Extinction through the Phanerozoic: A New Approach. The Journal of Geology 111:125–148.
- Foote, M., J. S. Crampton, A. G. Beu, and R. A. Cooper. 2008: On the bidirectional relationship between geographic range and taxonomic duration. Paleobiology 34:421–433.
- Goldberg, E. E., and J. Foo. Memory in trait macroevolution. *bioRxiv* (2018): 465971.
- Hagen, O., K. Hartmann, M. Steel, and T. Stadler. 2015: Age-dependent speciation can explain the shape of empirical phylogenies. Systematic Biology 64:432–440.
- Hagen, O., T. Andermann, T. B. Quental, A. Antonelli, and D. Silvestro. 2017: Estimating Age-Dependent Extinction: Contrasting Evidence from Fossils and Phylogenies. Systematic Biology 67:458–474.
- Jablonski, D. 2008: Species Selection: Theory and Data. Annual Review of Ecology, Evolution, and Systematics 39:501–524.
- ———. 2017: Approaches to Macroevolution: 1. General Concepts and Origin of Variation. Evolutionary Biology 44:427–450.
- Jones, D. S., and D. Nicol. 2016: Origination, survivorship, and extinction of rudist taxa. Journal of Paleontology 60:107–115.
- Liow, L. H., and J. D. Nichols. 2010: Estimating rates and probabilities of origination and extinction using taxonomic occurrence data: capture-recapture approaches BT - Quantitative Methods in Paleobiology. Quantitative Methods in Paleobiology 16:81–94.
- Liow, L. H., L. Van Valen, and N. C. Stenseth. 2011: Red Queen: From populations to taxa and communities. Trends in Ecology and Evolution 26:349–358.
- Liow, L. H., Reitan, T. and P. G. Harnik. 2015: Ecological interactions on macroevolutionary time scales: clams and brachiopods are more than ships that pass in the night. Ecology letters 18:1030-1039
- Lu, P. J., M. Yogo, and C. R. Marshall. 2006: Phanerozoic marine biodiversity dynamics in light of the incompleteness of the fossil record. Proceedings of the National Academy of Sciences 103:2736– 2739.
- Marshall, C. R. 2017: Five palaeobiological laws needed to understand the evolution of the living biota. Nature Ecology and Evolution 1:1–6.
- May, M. R., S. Höhna, and B. R. Moore. 2016: A Bayesian approach for detecting the impact of mass-

extinction events on molecular phylogenies when rates of lineage diversification may vary. Methods in Ecology and Evolution 7:947–959

- Morlon, H., T. L. Parsons, and J. B. Plotkin. 2011: Reconciling molecular phylogenies with the fossil record. Proceedings of the National Academy of Sciences 108:16327–16332.
- Nee, S. 2006: Birth-Death Models in Macroevolution. Annual Review of Ecology, Evolution, and Systematics 37:1–17.
- Parker, W. C., and A. J. Arnold. 2007: Species Survivorship in the Cenozoic Planktonic Foraminifera: A Test of Exponential and Weibull Models. Palaios 12:3.
- Payne, J. L., and S. Finnegan. 2007: The effect of geographic range on extinction risk during background and mass extinction. Proceedings of the National Academy of Sciences 104:10506–10511.
- Pearson, P. N. 1995: Investigating age-dependency of species extinction rates using dynamic survivorship analysis. Historical Biology 10:119–136.
- Pires, M. M., D. Silvestro, and T. B. Quental. 2017: Interactions within and between clades shaped the diversification of terrestrial carnivores. Evolution, 71:1855-1864.
- Rabosky, D. L. 2006: Likelihood Methods for Detecting Temporal Shifts in Diversification Rates. Evolution 60:1152.
- Raup, D. M. 1975: Taxonomic survivorship curves and Van Valen's Law. Paleobiology 1:82-96.
- Raup, D. M., and J. John Sepkoski. 1986: Periodic extinction of families and genera. Science 231:833-836.
- Raup, D. M., and N. Winter. 1978: Cohort Analysis of Generic Survivorship Cohort analysis of generic survivorship. Analysis 4:1–15.
- Roy, K. 1996: Higher taxa in biodiversity studies: Patterns from eastern Pacific marine molluscs. Philosophical Transactions of the Royal Society B: Biological Sciences 351:1605–1613.
- Sepkoski, J. J. 1975: Stratigraphic biases in the analysis of taxonomic survivorship. Paleobiology 1:343– 355
- Sepkoski, J. J. 1992: Phylogenetic and ecologic patterns in the Phanerozoic history of marine biodiversity. Pp 77-100. In: Eldredge, N., ed. Systematics, ecology and the biodiversity crisis. Columbia University Press: New York. NY.
- Sepkoski, J. J., McKinney, F. K., and S. Lidgard. 2000: Competitive displacement among post-Paleozoic cyclostome and cheilostome bryozoans. Paleobiology 26:7-18.

- Silvestro, D., J. Schnitzler, L. H. Liow, A. Antonelli, and N. Salamin. 2014a: Bayesian estimation of speciation and extinction from incomplete fossil occurrence data. Systematic Biology 63:349–367.
- Silvestro, D., N. Salamin, and J. Schnitzler. 2014b: PyRate: A new program to estimate speciation and extinction rates from incomplete fossil data. Methods in Ecology and Evolution 5:1126–1131.
- Silvestro, D., A. Antonelli, N. Salamin, and X. Meyer. 2018: Improved estimation of macroevolutionary rates from fossil data using a Bayesian framework. BioRxiv 18:316992.
- Silvestro, D., A. Antonelli, N. Salamin, T. Quental. 2015: The role of clade competition in the diversification of North American canids. Proceedings of the National Academy of Sciences 112:8684-8689.
- Töpel, M., A. Zizka, M. F. Calió, R. Scharn, D. Silvestro, and A. Antonelli. 2017: SpeciesGeoCoder: Fast categorization of species occurrences for analyses of biodiversity, biogeography, ecology, and evolution. Systematic Biology 66:145–151.
- Van Valen, L. 1973: A New Evolutionary Law. Evolutionary Theory 1:1-30.
- Žliobaitė, I., M. Fortelius, and N. C. Stenseth. 2017: Reconciling taxon senescence with the Red Queen's hypothesis. Nature 552:92–95.

Supplementary material

Supplementary file 1

from "Re-evaluation of the "law of constant extinction" for ruminants at different taxonomical scales"

The decision about the time window for the age-dependency analysis

The Diversification process commonly present shifts of speciation and extinction rates through time. As the duration of lineages is directly dependent on the extinction rate (Marshall, 2017), any shifts in extinction rate will affect the duration of lineages (Supplementary figure 2). To avoid the influence of this effect, we restricted our analyses to intervals of time showing constant extinction rate. We also excluded from analysis all taxa which had at least one occurrence crossing the limits of a given window (see below). All this information came from Januario and Quental (in prep., chapter 1) which recently described the temporal dynamics of the Ruminants under alternative taxonomies assumed over the integration of the Paleobiology Database (PBDB - <<u>http://paleobiodb.org/>; <http://fossilworks.org/></u>) and the New and Old Worlds Database (NOW - <<u>http://www.helsinki.fi/science/now/</u>>). Here, we used the taxonomy compiled by the Januario e Quental (in prep., chapter 1 - in the terminology of their article, we used the "OUR" database). The details fo the taxonomy choices made are described in Januario and Quental (in prep., chapter 1). We did not use another taxonomy described by those authors because we believe it to be the best taxonomic treatment of the date.

The authors also described the diversification dynamics for the ruminants in the species and at the genus level. The dynamics reported by the authors is of considerable complexity and the position of the extinction rate shifts varies according to the taxonomic level and the level of significance for the detection of the extinction rate shift (Supplementary figures 3 and 4). Below, we detail the choices we made for the delimitation of the time window of the age-dependency analysis we implemented, which takes in consideration the complexity of the time variation in extinction rates and the taxonomy level.

Species-level

If the significance considered is of a Bayes factor = 2, the dynamics described by Januario and Quental (in prep., chapter 1) in species level have three extinction rate shifts: one between 42 and 39.5 mya, one between 8 and 4 mya, and one between 1.5 and 0.5 mya. If the significance considered is of a Bayes factor = 6, then other tree shifts are recovered: \sim 6 mya, \sim 5 mya, and \sim 1 mya.

Because of these complex diversification scenario, we choose to run two analyses for the species dataset. One conservative, in which we used time intervals defined by extincion rate shifts with Bayes factor > 2 (at least "low significance" shifts), and a sensibility analysis in which we only used time intervals defined by extinction rate shifts with Bayes factor > 6

("highly significant" shifts). We used the 39.5 - 8 mya time window for the conservative analysis, and the interval between the root age of the clade (estimated through PyRate) until 6.5 mya time window for the sensibility analysis in the species level (Supplementary figure 3). Our final dataset for species level was composed of 1244 occurrences of 287 species in the conservative analysis (Supplementary figure 5), and 1704 occurrences for 361 species (Supplementary figure 6) in the sensibility analysis

Genera-level

If the significance considered is of a Bayes factor = 2, the dynamics described by Januario and Quental (in prep., chapter 1) at the genus level have three extinction rate shifts: one between 10 and 7 mya, one between 6.5 and 4 mya, and one between 1.5 and 0.5 mya. If the significance considered is of a Bayes factor = 6, then other four shifts are recovered: ~9 mya, ~6 mya, ~5 mya, and ~1 mya.

The time positions for the extinction rate shifts were similar under the "low significance" (Bayes factor = 2) and "high significance" (Bayes factor = 6), so we choose to run only one analysis considering all shifts. Hence we used the interval between the root age of the clade (estimated thought PyRate) until 10.5 mya time window for the analysis in the genus level (Supplementary figure 4). Our final dataset for the genera level was composed of 767 occurrences of 82 genera in (Supplementary figure 7).

References:

- Januario, M. L. And T. B. Quental (in prep. Chapter 1). Ruminantia macroevolutionary patterns are robust to distinct taxonomic treatments on data from paleontological online databases.
- Marshall, C. R. 2017: Five palaeobiological laws needed to understand the evolution of the living biota. Nature Ecology and Evolution 1:1–6.

Supplementary figures



Supplementary figure 1: The behaviour of the Weibull Distribution of lineage durations (panel A) and the extinction rate in function of age (panel B) according to the Hagen *et al* (2018) model. When lineages are under negative age-dependent extinction, the shape parameter of the expected distribution of lineages is smaller than 1 (panel A, dashed gray line) and the extinction probability is higher for young lineages, but decay as the lineage lives (panel B, dashed gray line). When lineages are under positive age-dependent extinction probability is lower for young lineages and increases as the lineage persists (panel B, continuous gray line). When lineages are under age-independent extinction, the shape parameter of the expected distribution of lineages is close to 1 (panel A, continuous black line) and the extinction remain constant through all lineage duration (panel B, continuous black line).



Supplementary figure 2: Different diversification regimes and their patterns of lineage duration. As lineage duration (black lines), at the species-level, only responds to variation in extinction rate (red lines), the variation in speciation rate (blue lines) does not affect lineage duration. We restricted our analysis in times were the extinction rate is constant (green hatched areas) and removed lineages which occurrences range overcome these moments (grey lines).



Supplementary figure 3. Rates through time (RTT) plots with HPD 95% for the posterior of extinction rate estimated using the "mG + qShift" preservation model in PyRate for the species dataset. The continuous line represents the median of the posterior distribution in each interval. The bottom lines represent low-significance times of rate shift, and the squares represent the times of highly significant rate shifts. The grey bars indicate the qShift intervals used in the analysis, and the green bars indicate the size of the time window used in each analysis.

Genera level analysis



Supplementary figure 4. Rates through time (RTT) plots with HPD 95% for the posterior of extinction rate estimated using the "mG + qShift" preservation model in PyRate for the genera dataset. The continuous line represents the median of the posterior distribution in each interval. The lines on the graph bottom represent low-significance times of rate shift, and the squares represent the times of highly significant rate shifts. The points in time where white and grey bars meet indicate the time points determining qShift intervals used in the analysis, and the green bar at the top of the graph indicates the size of the time window used in the analysis. Origination rate dynamics at the genus-level also shows constant rates (see chapter 1).



Species - Conservative Analysis

Supplementary figure 5. Occurrences used in the conservative analysis in the species level. The red lines represent occurrences of species totally present within the analysis window (and so in all PyRate replicas), and the gray lines represent occurrences of species which have potential to be sampled out of the window and so being excluded from analysis.

Species - Wider window Analysis



Supplementary figure 6. Occurrences used in the wider window analysis in the species level. The red lines represent occurrences of species totally present within the analysis window (and so in all PyRate replicas), and the gray lines represent occurrences of species which have potential to be sampled out of the window and so being excluded from analysis.



Genera Analysis

Supplementary figure 7. Occurrences used in the genus level analysis. The red lines represent occurrences of species totally present within the analysis window (and so in all PyRate replicas), and the gray lines represent occurrences of species which have potential to be sampled out of the window and so being excluded from analysis.

analysis.

Species level - Conservative



Role 🛱 ADE model 🛱 Clade origin 🛱 MCMC 🛱 preservation

Supplementary figure 8. ESS values for the PyRate parameters for all replicates in the conservative analysis in the species level. The blue parameters have a role in the general MCMC PyRate model. The crimson boxes refer to preservation parameters. The green box refer to the estimative of the Ruminantia origin and the orange boxes refer to parameters associated with the age-dependency weibull model, which also considers the species not sampled in the fossil record. We consider as suficiently sampled all replicas which ESS is larger than 200 (dashed line) for a given parameter. Note that 2000 is the maximum value the ESS can have for a given replica in our case.



Role 🛱 ADE model 🛱 Clade origin 🛱 MCMC 🛱 preservation

Supplementary figure 9. ESS values for the PyRate parameters for all replicates in the wider window analysis in the species level. The blue parameters have a role in the general MCMC PyRate model. The crimson boxes refer to preservation parameters. The green box refer to the estimative of the Ruminantia origin and the orange boxes refer to parameters associated with the age-dependency weibull model, which also considers the species not sampled in the fossil record. We consider as suficiently sampled all replicas which ESS is larger than 200 (dashed line) for a given parameter. Note that 2000 is the maximum value the ESS can have for a given replica in our case.



Role 🛱 ADE model 🛱 Clade origin 🛱 MCMC 🛱 preservation

Supplementary figure 10. ESS values for the PyRate parameters for all replicates in the genera level analysis. The blue parameters have a role in the general MCMC PyRate model. The crimson boxes refer to preservation parameters. The green box refer to the estimative of the Ruminantia origin and the orange boxes refer to parameters associated with the age-dependency weibull model, which also considers the species not sampled in the fossil record. We consider as suficiently sampled all replicas which ESS is larger than 200 (not shown) for a given parameter.



Supplementary figure 11. Shape parameter estimates for resampled iterations within each replica (ordenend by median value) in the species level (conservative) analysis. The red line indicates the age-independent pattern (shape value =).



Supplementary figure 12. Shape parameter estimates for resampled iterations within each replica (ordenend by median value) in the species level (wider window) analysis. The red line indicates the age-independent pattern (shape value =).



Supplementary figure 13. Shape parameter estimates for resampled iterations within each replica (ordenend by median value) in the genera level analysis. The red line indicates the age-independent pattern (shape value = 1).



Supplementary figure 14. Relationship between shape estimate and the estimate of preservation rate for each iteration for the interval between the root age of Ruminantia (estimated from PyRate) and 33.9 Ma in the genera level analysis. The dashed line represents a value of shape which indicates age-independency.



Supplementary figure 15. Relationship between shape estimate and the estimate of preservation rate for each iteration for the 33.9 - 23.03 Ma interval in the genera level analysis. The dashed line represents a value of shape which indicates age-independency.



Supplementary figure 16. Relationship between shape estimate and the estimate of preservation rate for each iteration for the 23.03 - 15.97 Ma interval in the genera level analysis. The dashed line represents a value of shape which indicates age-independency.



Supplementary figure 17. Relationship between shape estimate and the estimate of preservation rate for each iteration for the 15.97 - 11.65 Ma interval in the genera level analysis. The dashed line represents a value of shape which indicates age-independency.



Supplementary figure 18. Relationship between shape estimate and the estimate of preservation rate for each iteration for the 11.65 - 10.5 Ma interval in the genera level analysis. The dashed line represents a value of shape which indicates age-independency.

Conclusão geral

"As espécies e os grupos de espécies gradualmente desaparecem, uma após a outra, primeiro de um ponto, depois de outro, e finalmente do mundo"

Charles Darwin (1859) [tradução livre]

"Ignorar a extinção é provavelmente tão imprudente para um biólogo evolutivo quanto é para um demógrafo ignorar a mortalidade"

David Raup (1994) [tradução livre]

O que o registro fóssil pode nos dizer sobre a dinâmica macroevolutiva da vida na Terra e a duração das linhagens do planeta? Esta dissertação tentou usar o grupo dos ruminantes para explorar essa questão valendo-se de métodos na fronteira do conhecimento. Mais especificamente foram estimadas a variação temporal nas taxas de especiação e extinção, e se a probabilidade de extinção está associada à longevidade das linhagens levando em consideração as incertezas inerentes ao uso do registro fóssil. O registro fossil é a mais valiosa de todas as fontes de evidências para se estudar a extinção. Essa importância se deve ao fato do registro fóssil documentar diretamente possíveis momentos de origem e extinção das linhagens. Apesar de seu valor inerente, é absolutamente necessário que os padrões de diversificação biológica indicados pelo registro fóssil sejam analisados levando em consideração o fato de que a fossilização é um processo que gera uma miríade de vieses ou ruídos nos dados. A leitura direta do registro fóssil pode sugerir por exemplo que: (1) a primeira e última ocorrência de uma dada linhagem são de fato o momento de surgimento e extinção da mesma, quando isso é extremamente improvável; (2) poucas espécies existiram (i.e. duraram) por pouco tempo, quando todas as evidências apontam que essas espécies com longevidade muito pequena em geral não aparecem no registro fóssil. Além desses aspectos, ao se analisar o registro fóssil é preciso levar em consideração a possibilidade das taxas de preservação fóssil se alteram ao longo do tempo geológico, e que mudanças temporais nas taxas de extinção influenciam a duração das linhagens. O método e a delimitação amostral desta dissertação levam todos esses aspectos em consideração durante suas estimativas, além de também considerarem outros pontos importantes, discutidos em cada um dos capítulos. As descobertas de cada capítulo estão sumarizadas a seguir.

No primeiro capítulo, investigamos se, e como, as estimativas das taxas de especiação e extinção se alteram quando os dados são submetidos a diferentes tratamentos taxonômicos. Também comparamos as dinâmicas da diversificação estimadas no nível de espécie e de gênero, para investigar se o último pode ser usado como um "proxy" para a dinâmica no nível de espécies, uma prática comum em paleontologia. Para tal utilizamos tratamentos de diferentes complexidades e comparamos esses tratamentos com um tratamento feito por especialistas em

paleontologia de ruminantes. Encontramos, conforme esperado, que a curadoria nos dados influencia os resultados obtidos mas que, surpreendentemente, revisões guiadas pela informação do próprio banco de dados e implementadas com códigos automatizados já alcançam resultados similares a revisões muito mais detalhadas realizadas pelo mesmo indivíduo (banco de dados OUR, capítulo 1). Por outro lado, revisões detalhadas realizadas por diferentes indivíduos (um grupo de especialistas em ruminantes e um não-especialista), podem resultar em diferenças importantes na dinâmica de diversificação. Essas diferenças parecem estar diretamente relacionadas a aspectos da qualidade do registro fóssil das bases de dados, como por exemplo diferenças na resolução temporal das ocorrências entre as duas bases de dados, mas não podemos descartar a possibilidade que essas diferenças estejam relacionadas a opiniões taxonômicas distintas (apesar de haver pouca evidência que sugere isso). A comparação entre as análises nos níveis de gênero e espécie sugerem que dependendo do banco de dados a dinâmica macroevolutiva dos gêneros pode ser ou não uma boa aproximação da dinâmica das espécies.

No segundo capítulo, investigamos se a probabilidade de extinção das linhagens de ruminantes é independente da idade das linhagens, o que é previsto pela "lei das taxas constantes de extinção" proposta por Van Valen em 1973. Repetimos todas as análises em níveis de gênero e espécie. Encontramos que as espécies de ruminantes tendem a apresentar um consistente padrão de dependência de idade, com as espécies mais novas apresentando maiores probabilidades de se extinguirem quando comparadas com espécies com maiores durações. Entretanto, o regime que descreve a relação entre idade e probabilidade de extinção no nível dos gêneros sugerem que a probabilidade de extinção é independente da idade, apesar do ajuste dos dados ao modelo incluir uma grande incerteza, o que torna esse resultado menos contundente. Também encontramos associações fraca (nível de espécie) ou ausente (nível de gênero) entre a duração das linhagens e área de distribuição, e uma associação entre duração das linhagens e riqueza (para gêneros). Entretanto, nossos resultados apontam na direção de que podem existir outros fatores desconhecidos que regulam a duração das linhagens, tanto no nível de espécie quanto no nível de gênero, já que muita da variação na duração das linhagens não é relacionada nem à área de distribuição nem à riqueza de espécies.

O reconhecimento de que os padrões macroevolutivos podem ser resultado da natureza hierárquica e aninhada da organização biológica é reconhecida direta ou indiretamente tanto em trabalhos relativamente antigos (Simpson, 1951 (p. 80); Raup 1978) quanto atuais (Jablonski, 2017; Slater 2018). A natureza é intrinsecamente aninhada em diversas escalas ecológicas, espaciais, temporais e possivelmente evolutivas, e isso é algo que deve ser incorporado em nossas descrições sobre a história da vida. Em sua palestra, Simon Levin (1992) nos dá uma importante dica do que fazer quanto a esse aspecto: "a chave para entender como a informação é transferida entre escalas é determinar que informação é preservada e que informação é perdida conforme se move de uma escala para outra". De forma geral os resultados apresentados nos dois capítulos sugerem que a dinâmica das taxas temporais de especiação e extinção, além do regime de extinção dependente de idade descrita no nível dos gêneros nem sempre pode ser utilizada como "proxy" para a dinâmica presente no nível das espécies. Nesta dissertação, espero ter

contribuído de alguma forma a respeito dessa questão em particular, e certamente permanecerei atento a ela no futuro.

Referências

Darwin, C. 1859. On the origin of species by means of natural selection. Murray, London.

Jablonski, D. 2008: Species Selection: Theory and Data. Annual Review of Ecology, Evolution, and Systematics 39:501–524.

------. 2017: Approaches to Macroevolution: 1. General Concepts and Origin of Variation. Evolutionary Biology 44:427–450.

- Levin, S. A. 1992. The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. Ecology 73: 1943-1967.
- Marshall, C. R. 2017: Five palaeobiological laws needed to understand the evolution of the living biota. Nature Ecology and Evolution 1:1–6.
- Quental, T. B., and C. R. Marshall. 2010: Diversity dynamics: Molecular phylogenies need the fossil record. Trends in Ecology & Evolution 25:434–441.
- Rabosky, D. L. 2010. Extinction rates should not be estimated from molecular phylogenies. *Evolution: International Journal of Organic Evolution* 64:1816-1824.

Raup, D. M., and N. Winter. 1978: Cohort Analysis of Generic Survivorship Cohort analysis of generic survivorship. Analysis 4:1–15. Slater, Graham J., and A. R. Friscia. 2019. Hierarchy in adaptive radiation: A case study using the Carnivora (Mammalia). Evolution 73: 524-539.

Raup, D. M. 1994. The role of extinction in evolution. *Proceedings of the National Academy of Sciences* 91:6758-6763.

Simpson, G. G. 1951. The meaning of evolution. 192 pp. Yale University Press: New Haven, C.T.

Simpson, G. G. 1952. How many species? Evolution 6:342-342.

Slater, Graham J., and A. R. Friscia. 2019. Hierarchy in adaptive radiation: A case study using the Carnivora (Mammalia). Evolution 73: 524-539.

Van Valen, L. 1973: A New Evolutionary Law. Evolutionary Theory 1:1-30.