



**Universidade Federal do Rio Grande do Sul**

**Ciências Biológicas - Bacharelado**

**MATHEUS HEINRICH DRUMM**

Trabalho de Conclusão de Curso

**Ferramenta em R para análise da evolução de atributos ecológicos**

**TCC formatado de acordo com as normas da revista *Bioinformatics*, no**

**formato *Application Notes***

**Porto Alegre**

**Janeiro de 2018**

## RESUMO

**Sumário:** *Niche Shift Rate* (NSR) é uma função para o software R que calcula a mudança de um atributo fenotípico ao longo do tempo evolutivo juntamente com a direção dessa mudança. A função foi construída a partir de uma reconstrução de caracteres ancestrais (*ancestral character reconstruction* ou ACR) com o intuito de tornar a análise dos resultados da ACR mais clara e facilitar a interpretação. A função possui opção de utilizar uma única árvore filogenética ou múltiplas.

**Disponibilidade e implementação:** NSR é uma função para o ambiente R, publicada sobre uma licença de código aberto e gratuitamente disponibilizado na plataforma GitHub (<https://github.com/MHDrumm/nsr.R>)

**Contato:** [matheushdrumm@gmail.com](mailto:matheushdrumm@gmail.com)

## 1 Introdução

Compreender como se deu a evolução do nicho ou de um atributo de uma linhagem é um desafio para a ecologia. Diferentes linhagens de seres vivos demonstram uma grande diferença em suas taxas de evolução fenotípica (Adams, 2014). Tais taxas podem ser expressas de inúmeras maneiras, uma delas é o desvio padrão fenotípico por geração (Haldane, 1949). Essas diferenças de taxa de evolução fenotípica podem ser atribuídas a diversos motivos. Por exemplo, variações ecológicas entre linhagens podem determinar diferenças nas taxas evolutivas em atributos fenotípicos (Collar et al. 2009; Mahler et al. 2010). Além disso, as diferenças nas taxas de evolução fenotípica frequentemente se expressam durante períodos de radiação adaptativa (Cooney et al. 2017). Por sua vez, a teoria de nicho prediz que para uma espécie persistir em um local ela deve ter um conjunto particular de atributos que permitam ela superar os filtros ambientais e coexistir com as outras espécies (Keddy & Weiher, 1999).

Porém, não podemos analisar os atributos atuais das espécies de uma dada linhagem e esperar que eles respondam todas as nossas perguntas ecológicas. Precisamos examinar a trajetória evolutiva dessa linhagem para entender os processos que deram origem às espécies que observamos atualmente.

Com o crescimento das áreas de ecologia filogenética e funcional, as informações filogenéticas (Chase et al. 2016; Jetz et al. 2012) e de atributos (Kattge et al. 2011; Wilman et al. 2014) estão tornando-se mais disponíveis. Devido a isso, novos métodos e ferramentas que integrem informação filogenética e funcional são essenciais ao desenvolvimento do conhecimento ecológico em larga escala, pois permitem a análise de grandes quantidades de dados e interpretação de padrões que ultrapassam o que poderia ser observado em campo.

NSR é uma função para o software R (R Core Team, 2017). Publicada sob licença de código aberto, gratuitamente disponível na plataforma GitHub (<https://github.com/MHDrumm/nsr.R>) que implementa um método para calcular o quanto um atributo de uma espécie mudou ao longo do tempo evolutivo da linhagem a partir de uma reconstrução de caracteres ancestrais (ACR). Usualmente os resultados de uma ACR são apresentados na forma de uma filogenia com um gradiente de cores expressando a mudança nos atributos. Como podemos ver na figura 1A existe um padrão de mudança de cor que coincide nas duas ACR para atributos diferentes. Porém essa abordagem é propensa a gerar falhas, já que a interpretação é feita visualmente observando as cores da filogenia, que podem ser confusas dependendo de quantas espécies ela contém. Já o NSR permite mensurar a mudança do atributo durante a evolução da linhagem juntamente com o sentido dessa mudança no formato de um índice (Tamanho de efeito padronizado), por exemplo uma espécie que ficou muito maior teria um valor elevado de NSR, utilizando somente uma filogenia e os valores de atributos atuais das espécies.

## 2 Descrição

### 2.1 Manipulação de dados

A função NSR utiliza a mesma manipulação de dados usualmente utilizada no software R. Para realizar a análise, dois conjuntos de dados são necessários: vetor nomeado contendo os valores de atributos das espécies e uma árvore filogenética ultramétrica (ou um conjunto de filogenias ultramétricas). Os atributos precisam ser descritos com valores numéricos contínuos e estarem com os mesmos nomes que as espécies na filogenia. O arquivo da árvore pode ser tanto um arquivo *phylo* quanto um *multiPhylo*, sendo este último mais recomendado pois possibilita um maior número de opções de análises posteriores.

### 2.2 Cálculo do Índice

Para calcular o índice NSR, a função primeiro realiza uma reconstrução de caracteres ancestrais de máxima verossimilhança (ML *ancestral character reconstruction*), utilizando a função ‘fastAnc’ do pacote *phytools* (Revell, 2012). Esta função reconstrói o valor de atributo para cada nó da filogenia com base nos atributos atuais das espécies. Logo após, os nós que compõem o ramo conectando cada espécie até o nó raiz da filogenia são mapeados. Com a informação de quais nós compõem a trajetória evolutiva de cada espécie e os valores de atributo reconstruído para cada nó, é calculado o índice de taxa de mudança do nicho (*Niche Shift Rate*) para cada espécie da árvore usando a seguinte fórmula:

$$NSR = \frac{X_A - \bar{X}_R}{dp\bar{X}_R}$$

Onde  $X_A$  é o atributo atual da espécie,  $\bar{X}_R$  é a média dos atributos reconstruídos correspondentes a cada nó ao longo da trajetória evolutiva daquela espécie, e  $dp\bar{X}_R$  é o desvio padrão dessa média. Valores iguais a zero indicam que o atributo atual não se alterou em

relação ao valor médio do atributo reconstruído até raiz da árvore. Já valores positivos ou negativos mostram um aumento ou decréscimo, respectivamente, do valor do atributo em relação ao valor médio do atributo reconstruído. Quando a árvore de entrada é um arquivo *phylo*, a função retorna o valor de índice para cada espécie. Para um arquivo *multiPhylo*, ela retorna um valor de índice por espécie para cada árvore dentro do arquivo.

### 2.3 Argumentos opcionais para arquivos *multiPhylo*

Para arquivos *multiPhylo* a função tem dois argumentos opcionais. O argumento *mean* quando verdadeiro muda o retorno da função para a média dos índices de cada árvore. Já o argumento *bootstrap*, quando é verdadeiro calcula o valor médio e os intervalos de confiança superiores e inferiores pelo procedimento de *bootstrap*, e adiciona esses valores ao retorno da função.

## 3 Conclusão

O método implementado na função NSR para R permite a integração de informação ecológica - na forma de atributos - e da história evolutiva da linhagem - na forma da filogenia. Dessa maneira, pode-se *entender como os atributos mudaram ao longo da evolução*. Uma das vantagens do NSR é que ele utiliza os atributos atuais das espécies, uma informação relativamente simples de conseguir e que cada vez mais é disponibilizada em bancos de dados. Além disso, o NSR expressa um índice numérico para cada espécie, diferente do ACR no qual ele se baseia. A figura 1A apresenta a reconstrução de dois atributos para aves da família Tyrannidae, onde podemos observar um padrão visual que torna a análise pouco interpretável. Já na Figura 1B, temos um modelo linear entre o NSR dos mesmos atributos da Figura 1A. Com o modelo linear podemos testar a relação da mudança dos dois atributos e como o NSR gera um valor para cada espécie outros processos podem ser realizados, por exemplo, esses valores podem ser plotados em um mapa para entendermos como essas mudanças se estruturam

no espaço. Como podemos ver a Figura 1B apresenta uma abordagem analítica mais objetiva e robusta, pois além do padrão visual, o modelo linear permite calcular um ajuste do modelo ( $R^2$ ) e uma probabilidade de rejeição da  $H_0$ . Além disso, a implementação de métodos novos no software R é o ideal para o desenvolvimento do nosso conhecimento ecológico como um todo, já que uma vez publicados podem ser facilmente modificados para se ajustar às necessidades de diferentes pesquisadores.

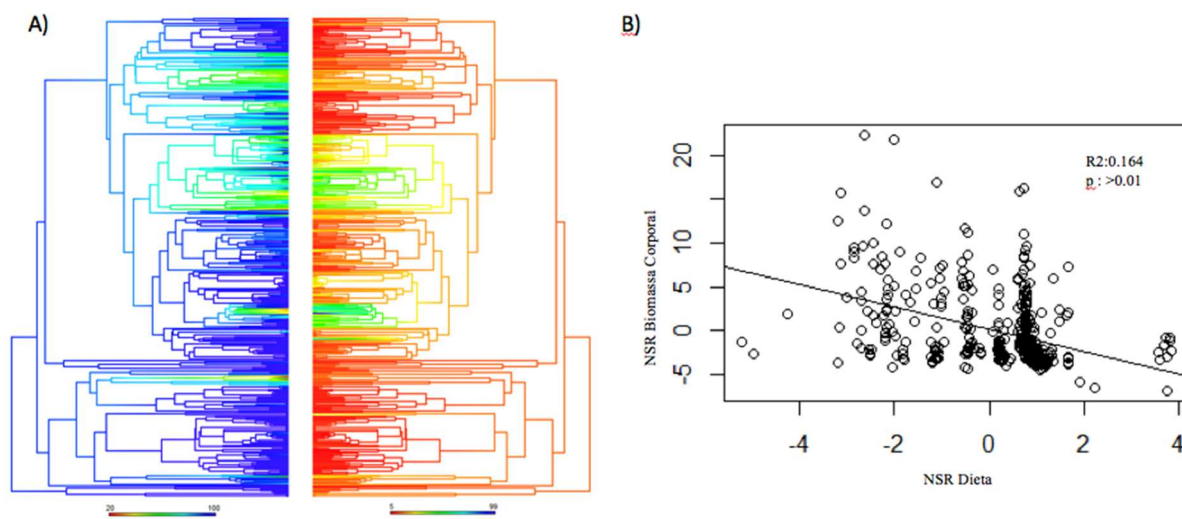


Figura 1: A) ACR de dieta (esquerda) e biomassa corporal (direita) das aves da família Tyrannidae. Na ACR de dieta os valores estão expressos em % de insetivoria e na ACR de peso em gramas. B) Modelo linear entre o NSR utilizando o atributo biomassa corporal e dieta, que estão expressos da mesma maneira que nas ACRs da figura 1A. O modelo os valores de NSR utilizados foram a média de 100 árvores.

## Agradecimentos

Agradeço à Lucas Porto por ceder um conjunto de dados para testes e toda equipe do Laboratório de Ecologia Filogenética e Funcional da UFRGS pelos conselhos.

Financiamento: Matheus H. Drumm recebeu financiamento do Conselho Nacional de Desenvolvimento Tecnológico e Científico (CNPQ).

Conflito de Interesse: Nenhum declarado.

## Referências

Adams, D. C. (2014) Quantifying and comparing phylogenetic evolutionary rates for shape and other high-dimensional phenotypic data. *Syst. Biol.*, 63, 166–177.

Chase, M. W. et al. (2016) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.*, 181, 1–20.

Collar, D.C., O'Meara, B.C., Wainwright, P.C. & Near, T.J. (2009). Piscivory limits diversification of feeding morphology in centrarchid fishes. *Evolution*, 63, 1557–1573.

Cooney, C. R. et al. (2017) Mega-evolutionary dynamics of the adaptive radiation of birds. *Nature*, 542, 344–347.

Haldane, J. B. S. (1949). Suggestions as to quantitative measurement of rates of evolution. *Evolution*, 3, 51-56.

Jetz, W. et al. (2012) The global diversity of birds in space and time. *Nature*, 491, 444–448.

Kattge, J. et al. (2011) TRY - a global database of plant traits. *Glob. Chang. Biol.*, 17, 2905–2935.

Keddy, P. and Weiher, E. (1999) Introduction: the scope and goals of research on assembly rules. In Weiher, E. and Keddy, P. (eds.) *Ecological Assembly Rules: Perspectives, Advances, Retreats*. Cambridge University Press, Cambridge, UK, pp. 1–20.

Mahler, D.L., Revell, L.J., Glor, R.E. & Losos, J.B. (2010). Ecological opportunity and the rate of morphological evolution in the diversification of Greater Antillean anoles. *Evolution*, 64, 2731–2745.

R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Revell, L. J. (2012) phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3 217–223. doi:10.1111/j.2041-210X.2011.00169.x

Wilman, H. et al. (2014) EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. *Ecology*, 95, 2027–2027.