

## LACUNAS EM BANCOS DE DADOS VIRTUAIS E AS IMPLICAÇÕES PARA ESTUDOS BOTÂNICOS

Michel Biondi<sup>1x</sup>, Carolina Njaime Mendes<sup>2</sup> & Ravi Fernandes Mariano<sup>3</sup>

(<sup>1</sup>Universidade Federal de Lavras, Departamento de Biologia, Campus Universitário, Lavras, Minas Gerais, 37.200-000; <sup>2,3</sup>Universidade Federal de Lavras, Departamento de Ciências Florestais, Campus Universitário, Lavras, Minas Gerais, 37.200-000; <sup>x</sup>Autor de correspondência: [michel.biologicas@gmail.com](mailto:michel.biologicas@gmail.com))

### INTRODUÇÃO

A disponibilidade de dados online é um grande avanço para a ciência em diversas áreas do conhecimento (Canhos *et al.* 2015). Essas informações auxiliam pesquisadores em análises cada vez mais robustas, com maior abrangência e magnitude amostral (Canhos *et al.* 2015). Porém, nem sempre estes dados estão completos, por vezes podem haver ruídos ou lacunas não preenchidas já que não existe um padrão geral para sua disponibilização, mas sim uma proposição de sugestões para cada tipo de banco de dados (Peixoto *et al.* 2006). Bancos de dados mais restritivos acabam por vezes tendo a inclusão de informações vinculadas a publicações em periódicos, como o GenBank (Genbank 2018), outros são menos restritivos, porém exigem o formato em que os arquivos devem ser anexados, como exemplo temos o Species Link que tem seu banco de dados alimentado por diversos softwares, sendo o Brahms mais utilizado (Peixoto *et al.* 2006).

No entanto, as falhas encontradas, tal como dados faltantes sobre as fisionomias vegetais, pontos mal georreferenciados e material sem identificação (Cabanillas & Silva 2018), muitas vezes limitam as análises, restringindo o espectro de perguntas a serem respondidas (Peixoto *et al.* 2006). Contudo, é certa a evolução na qualidade dos dados disponibilizados, já que estes bancos de dados são revisados periodicamente e a tendência é que a utilização de dados online seja ampliada para a pesquisa científica (Cabanillas & Silva 2018).

Neste trabalho buscamos exemplificar com dados vegetacionais obtidos na base de dados do Species Link que as informações faltantes geram limitações as pesquisas básicas e as aplicações ecológicas são prejudicadas por dados erroneamente ajustados.

### MATERIAL E MÉTODOS

Foram obtidos dados de vegetação na plataforma digital *Species Link*, tendo como filtro o município de Lavras, cidade localizada na região do Campo das Vertentes no estado de Minas Gerais, onde encontra-se a Universidade Federal de Lavras. Nela está situado o herbário ESAL, parte importante nas pesquisas relacionadas à vegetação, possuindo um acervo de cerca de 30.000 exsicatas e recentemente com seus dados disponibilizados na internet, depois de um processo de informatização que se iniciou por volta dos anos 2000 (Massensini Júnior *et al.* 2017). Posteriormente foi montada uma planilha, onde foram selecionados os seguintes itens advindos do banco de dados: Família, gênero, espécie, coordenadas geográficas, localidade e anotações. A partir dessa planilha foram contabilizados os vouchers com informações completas e aqueles com dados faltantes.

### RESULTADOS E DISCUSSÃO

Foram obtidos 11.348 vouchers de exsicatas (tabela 1) relacionados a busca Lavras no sítio Species Link, destes o maior dado faltante foi o de georreferenciamento (89,35%), o que já era esperado visto que a tecnologia do georreferenciamento é relativamente recente e as coletas disponibilizadas muitas vezes são antigas, sendo talvez o maior problema para trabalhos com o viés de distribuição de espécies, já que sem saber o ponto exato de coleta, pesquisas em ecologia espacial e biogeografia através dos bancos de dados de vegetação acabam por ter uma amostragem diminuta, as aproximações que são feitas a posteriori são outro problema (Cabanillas & Silva 2018).

Esses ruídos são ocasionados quando colocamos por exemplo a localização do município e não a do local de coleta, sendo este tipo de atitude recorrente (Cabanillas & Silva 2018). Neste caso o problema dependerá da escala adotada no estudo. Estudos com escalas grandes e de ocorrência de espécies não apresentarão um grande problema, porém se a escala for menor, ou o estudo exigir uma distribuição mais refinada, os bancos de dados vegetacionais podem não ajudar.

Ainda sobre problemas relacionados a trabalhos de distribuição de espécies, podemos citar equívocos no preenchimento da localidade (19,97%), úteis para orientar e limitar a região estudada. As anotações de campo (16,46%), interessantes para obtermos informações acerca da fisionomia vegetal ou hábito de vida das plantas são mais detalhadas em coletas mais recentes e podem ajudar quando estamos trabalhando com uma dessas informações como objeto de estudo (Cabanillas & Silva 2018).

Um erro de preenchimento que atrapalha a busca em si é o município (2,28%). Muitos tem nomes parecidos e por si só já geram dúvidas (por exemplo, na busca por Lavras, encontramos coletas de Lavras Novas (Cabanillas & Silva 2018).

Por fim, vemos a importância de taxonomistas, já que 22,53% (2.557) das exsicatas não estão identificadas a nível específico e nesta busca em um município, encontramos 1,45% (164) de exsicatas sem identificação em nível de família. Neste estudo não avaliamos as identificações, portanto elas podem apresentar diversos sinônimos nomenclaturais e identificações equivocadas, o que aumenta a importância dos taxonomistas visitarem as coleções biológicas revisando-as (Peixoto & Morim 2003).

**Tabela 1.** Relação entre dados faltantes e completos encontrados na busca no sítio *Specieslink* com filtro para Lavras.

	Dados faltantes	Dados completos
<b>Família</b>	1,45% (164)	98,55% (11184)
<b>Município</b>	2,28% (259)	97,72% (11089)
<b>Gênero</b>	4,25% (482)	95,75% (10866)
<b>Anotações</b>	16,46% (1868)	83,54% (9480)
<b>Localidade</b>	19,97% (2266)	80,03% (9082)
<b>Espécie</b>	22,53% (2557)	77,48% (8791)
<b>Coordenada</b>	89,35% (10140)	10,65% (1208)

## CONCLUSÃO

Os bancos de dados online de livre acesso são fundamentais para a evolução da ciência e a manutenção destes com revisões periódicas são essenciais para o bom funcionamento e cumprimento de seu propósito. A geração e disponibilização de informação de qualidade para a exploração científica é de suma importância para a evolução das ciências de um modo geral. Para a botânica, vemos que ainda existem pontos a serem aprimorados e eles dependem de nós cientistas, desde o momento da coleta até o momento de incluirmos nossos dados nos herbários, para que posteriormente as análises sobre a biodiversidade utilizando os bancos de dados virtuais sejam cada vez mais abrangentes e descrevam a natureza e seus padrões com mais robustez.

## AGRADECIMENTOS

Agradecemos a CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pela concessão das bolsas de estudos que possibilitaram as pesquisas no decorrer do período do doutoramento.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Cabanillas C, Silva LA (2018) Aplicação da mineração de dados na melhoria da qualidade de dados em coleções botânicas. e-Revista Facitec, 1(11). Disponível: <http://periodicosbh.estacio.br/index.php/e-revistafacitec/article/view/4839>. Acessado em 25 de maio de 2018.
- Canhos DAL, Sousa-Baena MS, Souza S, Maia LC, Stehmann JR, Canhos VP, De Giovani R, Bonacelli MBM, Los W, Peterson AT (2015) The importance of biodiversity e-infrastructures for megadiverse countries. PlosBiology 13 (7): e1002204. Doi: 10.1371/journal.pbio.1002204
- GenBank. <https://www.ncbi.nlm.nih.gov/genbank/>. Acessado em 25 de maio de 2018.
- Massenssini Júnior A, Mourão B, Mansanares ME (2017) Extensão como instrumento de desenvolvimento socioambiental e de formação científica: as experiências do Herbário ESAL-UFLA, Lavras, Minas Gerais. Unisanta BioScience, 6(5): 172-176.
- IBGE. <https://www.ibge.gov.br/>. Acessado em 25 de maio de 2018.
- Peixoto AL, Morim MP (2003) Coleções botânicas: documentação da biodiversidade brasileira. Ciência e Cultura, 55(3): 21-24.
- Peixoto AL, Barbosa, MRDV, Menezes M, Maia LC, Vazoleler RF, Marinoni L, Canhos DAL (2006) Diretrizes e estratégias para a modernização de coleções biológicas brasileiras e a consolidação de sistemas integrados de informação sobre biodiversidade. Brasília: Centro de Gestão e Estudos Estratégicos: Ministério da Ciência e Tecnologia. 324p.
- Species Link. <http://slink.cria.org.br>. Acessado em 20 de maio de 2018.