



FLORA DE BOA NOVA: USO DE BANCOS DE DADOS ONLINE COMO IMPORTANTES FERRAMENTAS PARA LEVANTAMENTOS FLORÍSTICOS

ANA JULIA DE SOUZA SAMPAIO; ANDREA KARLA ALMEIDA DOS SANTOS

RESUMO

A Bahia é um dos estados brasileiros que possui grande biodiversidade, principalmente por sua grande extensão territorial e os diferentes domínios fitogeográficos (Cerrado, Mata Atlântica e Caatinga). A existência de UCs no estado é uma importante iniciativa na proteção dessa diversidade biológica, a exemplo do PARNA e REVIS de Boa Nova, localizado entre ambientes de transição de Caatinga e Mata Atlântica. Apesar de recorrentes coletas botânicas feitas na área nos últimos anos, ainda são poucos os trabalhos publicados sobre a diversidade florística da região que inclui o PARNA e o REVIS. O presente estudo teve como intuito buscar, filtrar e analisar dados advindos de bancos de dados online (SpeciesLink e REFLORA Herbário Virtual) para fazer um levantamento geral sobre a diversidade da flora da região, bem como fazer uma apuração através de revisão bibliográfica. A limpeza de dados foi feita com auxílio de ferramentas do próprio SpeciesLink, junto com informações advindas do Reflora e, por fim, uma curadoria manual das informações. Sistemas de informações como o SpeciesLink são úteis para pesquisa, educação e políticas públicas, mas podem conter erros e ambiguidades, levando a interpretações equivocadas na hora de usar os dados. O refinamento criterioso dos dados é necessário para obter informações mais seguras. Verificando e limpando os bancos de dados, encontramos um total de 2249 registros não duplicados para a região de Boa Nova, sendo que duas espécies são indicadas como Criticamente em Perigo (CR) e 25 classificadas como Em Perigo (EN). Além disso, também houve registros de espécies endêmicas de áreas de Cerrado, o que constitui um achado importante visto que ainda não na literatura estudos de áreas de Cerrado para Boa Nova, o que é muito significativo para o conhecimento da diversidade da região, até então reconhecida por abrigar espécies da Caatinga, Mata Atlântica e da típica Mata-de-Cipó (floresta estacional) do sudoeste da Bahia.

Palavras-chave: Sistemas de informações; SpeciesLink; Unidades de Conservação; Reflora; Cerrado.

1 INTRODUÇÃO

A Bahia destaca-se entre os estados brasileiros por sua notável variedade de ecossistemas, abriga uma rica diversidade de biomas e suas áreas de transição, proporcionando um ambiente propício para a existência de uma ampla variedade de espécies (SEMA/BA; INEMA, 2017). Na região sudoeste da Bahia, encontra-se o Parque Nacional (PARNA) e o Refúgio da Vida Silvestre (REVIS) de Boa Nova, essas unidades de conservação (UC) tem como objetivo proteger ambientes de transição entre a Caatinga e a Mata Atlântica, com ênfase na preservação da flora e fauna da região (BRASIL, 2010). A região de Boa Nova já foi alvo de estudos florísticos como nos trabalhos de Simões, Zappi, & Aona (2020); Fonseca *et al.*, (2020); Rêgo & Azevedo (2016) e Brandão (2014), além disso as coletas mais recentes são

feitas principalmente por professores (as) e pesquisadores (as) de universidades públicas (++speciesLink network, 2023). Os espécimes são depositados em herbários e a maioria dos dados a eles associados, alimenta os bancos de dados online como o SpeciesLink ou o Re flora. O SpeciesLink é um sistema de informações que surgiu em 2001 como iniciativa aprovada pela FAPESP, idealizado para facilitar o compartilhamento de dados biológicos coletados e dificultar a perda dos mesmos quando passados para o digital, tendo como objetivo integrar o seu antecessor (SinBiota) com registros de coleções biológicas de Universidades e Institutos de São Paulo. Na versão atual do SpeciesLink, a *e*-infraestrutura já conta com a participação de várias organizações e ao menos uma coleção de cada estado brasileiro (CANHOS *et al.*, 2022).

Apesar de conter uma grande quantidade de dados de diferentes, o SpeciesLink não tem a capacidade de filtrar os dados que recebe, ficando sob responsabilidade das coleções o controle da qualidade dos dados adicionados no banco de dados. Entretanto o Centro de Referência em Informação Ambiental (CRIA) auxilia na manutenção do site apoiando as coleções através de relatório de limpeza de dados, que ajuda curadores e gestores das coleções participantes a identificar erros e preencher informações ausentes nos espécimes biológicos, melhorando a qualidade dos dados no sistema, dessa forma se mantém uma mínima interferência na rotina das coleções com mecanismos que possibilitam a coleção ter controle sob seu compartilhamento de dados (CANHOS *et al.*, 2022).

Atualmente, o SpeciesLink conta com um aumento crescente nos registros para a região de Boa Nova, advindo principalmente de coletas feitas por universidades públicas (++speciesLink network, 2023) e, apesar dos mecanismos para verificação de erros, quando se faz uma consulta alguns persistem e só podem ser verificados numa minuciosa análise manual.

O presente trabalho teve como objetivo fazer um levantamento dos registros florísticos do município de Boa Nova disponíveis em coleções biológicas, através dos bancos de dados e bibliografia, além de realizar o tratamento e limpeza das informações obtidas e apresentar uma metodologia prática para filtragem desse tipo de dados.

2 METODOLOGIA

Utilizando os sites SpeciesLink e Re flora (++speciesLink network; Flora e Funga do Brasil) foi feito o levantamento dos registros botânicos marcados tanto para o Parque Nacional de Boa Nova quanto para Boa Nova-Ba, pois muitos registros na área do Parque não são cadastrados como fazendo partes da UC na hora de inserir os dados no site. Ou as vezes o contrário também ocorre. Além disso, como existem duas UC com fragmentos intercalados na região nem sempre é possível ter certeza se a amostra foi coletada no PARNA, no REVIS ou fora de ambas, assim o estudo está sendo feito para a região e não exclusivamente para área dentro das UC. Para baixar a lista foi utilizado os filtros de localização da coleta e tipo de coleção que está inserida conforme cada um dos sites de busca. Também foi feito o levantamento das espécies a partir da consulta as bibliografias disponíveis sobre as UC de Boa Nova, esse conjunto de dados foi compilado em uma Planilha do Google.

Além da busca através dos Sistemas de Informação online, também foi feito busca por levantamentos botânicos da área de Boa Nova na bibliografia. Os táxons encontrados foram acrescidos na planilha com os táxons vindos dos bancos de dados.

Após a obtenção dos dados se iniciou o processo de filtragem dos mesmos, muitos dos registros eram duplicatas, então o primeiro passo foi unir os dados que as diferentes fontes de dados tinham em comum, foi levado em conta o número de tombo dos registros e a identificação das espécies. O passo seguinte foi unir os registros de duplicatas que foram identificados pela mesma coleta através da aba de duplicatas do SpeciesLink. Posteriormente esse processo foi

refinado e feito manualmente, verificando nome e números de coletores um a um pois nem todas as duplicatas são registradas como sendo da mesma coleta no site. Após a lista não ter mais dados repetidos em relação aos registros, foi feita uma análise visual entre os registros que não estavam identificados até o nível de espécies dentro do mesmo gênero, para aqueles que possuíam imagens e tinham características visuais que poderiam ser identificados, assim foram reunidos como sendo um mesmo táxon os registros que eram visualmente iguais ou como táxons diferentes registros que se tratavam de espécies distintas. Para os demais casos de espécimes do mesmo gênero que não possuíam imagens para comparar e verificar se tratava do mesmo táxon ou não, foi mantido como táxons diferentes.

Alguns problemas foram recorrentes nos bancos de dados, relacionados a diferentes identificações para duplicatas da mesma coleta, assim foi estabelecida uma série de critérios para seguir e tentar solucionar os problemas e discrepâncias, para decidir qual nome deveria ser mantido ou excluído da lista de táxons. Nosso critério principal foi confiar sempre no nome que foi identificado por um taxonomista especialista do grupo estudado:

I. Muitas vezes havia duas ou três duplicatas, cada uma com uma identificação diferente, neste caso, verificava-se qual destas identificações foi feita por especialista da família, e decidia-se por considerar a identificação do especialista a correta.

II. Quando mais de uma identificação diferente era feita por diferentes especialistas, contava-se como um único táxon, porém deixa-se na lista os dois nomes referidos.

III. Quando as diferentes identificações eram feitas por pessoas que não eram especialistas e não tinha imagens que a gente pudesse comparar com a bibliografia, eram mantidos os dois nomes na listagem, porém o táxon era contato apenas uma vez.

Para os casos onde o nome estava escrito de forma errada ele era corrigido para a escrita correta, quando o nome não era mais aceito se procurava o sinônimo aceito e correto através de consultas ao Re flora. Quando se haviam dúvidas sobre a existência do nome foi feito consultas no site International Plant Names Index (IPNI)

A partir da lista filtrada foi criada uma lista a parte somente com os táxons encontrados, com base nesta foi feita uma busca no Re flora táxon por táxon para coletar dados referentes aos domínios fitogeográficos onde se encontram e o grau de Conservação.

3 RESULTADOS E DISCUSSÃO

3.1 Diversidade florística em Boa Nova

Durante a revisão bibliográfica foi possível perceber que ainda são poucos os trabalhos envolvendo a flora das UCs de Boa Nova, sendo maior parte deles produzida através de projetos com parcerias com universidades públicas, como o levantamento das famílias Cactaceae (SIMÕES, ZAPPI, & AONA, 2020) e Rubiaceae (FONSECA *et al.*, 2020) feito em parceria com a UFRB e o levantamento dos lajedos de Boa Nova (BARBOSA *et al.*, 2023), da família Orchidaceae (RÊGO & AZEVEDO, 2016) e trepadeiras (BRANDÃO, 2014) feito em parceria com as Universidades UESB e UFBA. Ambas as universidades possuem herbários que estão entre os quatro com maior quantidade registros na área do Parque (SpeciesLink, 2023).

Usando os critérios de busca estabelecidos, antes de fazer a limpeza dos dados, foram encontrados 9996 registros no SpeciesLink e 1673 no Re flora. Após seguir todos os critérios de limpeza dos dados com a exclusão de dados repetidos, comparação das identificações e manutenção apenas das espécies diferentes a lista final de táxons resultou em 2249 registros (não duplicados) para Boa Nova.

Com a apuração e minuciosa filtragem dos dados botânicos disponíveis para o município de Boa Nova foram encontrados registros de 186 famílias, 716 gêneros e 1218

espécies. Das espécies encontradas 8 foram encontradas exclusivamente pelo IPNI e não pelo Re flora, haviam também 7 registros que não foram encontrados em ambos (IPNI e Re flora), destes 2 registros correspondiam a nomes que não foram encontradas as grafias corretas nem sinônimos, enquanto 5 foram encontrados em outros locais além do IPNI e do Re flora. Todos os trabalhos sobre a vegetação de Parque encontrados abordaram levantamento de grupos específicos, sendo eles: Orchidaceae; Cactaceae; Rubiaceae e trepadeiras. Das 1218 espécies encontradas, 47 foram encontradas exclusivamente em trabalhos publicados e não nos bancos de dados online.

Dentre os dados filtrados há o registro de duas espécies indicadas como Criticamente em Perigo (CR) para a área do Parque Nacional de Boa Nova, *Syagrus santosii* (Arecaceae) e *Neoregelia brownii* (Bromeliaceae), além de 25 espécies que se encontram classificadas como Em Perigo (EN).

Apesar de Boa Nova se encontrar num ecótono entre Mata Atlântica e Caatinga, há registro de 13 espécies endêmicas de Cerrado e outras 666 que também ocorrem no domínio fitogeográfico, mas não são exclusivas dele. Essa informação se torna interessante uma vez que na área há existência de locais com fitofisionomia semelhante ao Cerrado na região que ainda não foram registradas oficialmente. Os objetivos da criação do Parque Nacional de Boa Nova incluem a proteção dos ecossistemas de transição entre a Mata Atlântica e a Caatinga já conhecidos na região. No entanto, nos limites das UC, também encontramos uma área com predominância de Cerrado, uma vegetação mais aberta e campestre, que possui uma fitofisionomia única na região.

3.2 Protocolo e análise sobre os bancos de dados

Com relação a análise dos dados, entre as listas baixadas dos sistemas de informação havia muitos dados com erros de digitação, incongruências sobre coletor e número de coleta, junção de gêneros com famílias erradas e ainda fotos de espécies diferentes do que deveria ser da mesma planta. Esses tipos de inconformidades na hora de acrescentar as informações nos bancos de dados podem levar a problemas durante a coleta de informações, durante a interpretação dos dados e trabalho com os mesmos. Além dos erros citados também foram encontrados dados incompletos, como a falta de localização, determinadores, nome de coletor, número de coleta e número de tombo.

Outro problema recorrente foi a falta de padronização dos dados, como o banco é abastecido por diferentes pessoas e instituições há diferenças na forma como eles são descritos, a inconformidade mais comum era em relação ao nome do coletor, o sobrenome vir antes ou depois, o nome ser escrito sem abreviações ou com, e até o nome ser incluído junto com o número de coletor de forma repetida, todos esses fatores influenciaram no que os bancos de dados interpretam como registro sendo duplicata ou não.

Apesar da interface apresentar ferramentas que auxiliem a evitar erros e enganos a respeito dos dados, como o sistema de duplicatas, é necessário que se tenha o mesmo número de coleta, nome de coletor e data (CANHOS *et al.*, 2022). O que se observou foi que pequenas alterações na forma de escrita ou ordem dos dados faz com que a interface do SpeciesLink não interprete os dados como sendo duplicatas, apesar de serem.

Conforme relatado por Canhos *et al.*, (2022), a ferramenta para busca de duplicatas pode auxiliar a recuperação de registros por curadores de coleção, para que se possa fazer a atualização ou correção de erros de identificação.

Além disso, desde 2012 o CRIA produziu uma ferramenta que permite usuários a deixarem observações a respeito dos registros dentro do SpeciesLink, essas informações são repassadas aos curadores das coleções para verificação e atualização das informações, desde a criação desse mecanismo até 2022 foram enviadas aproximadamente 19 mil notas, a

colaboração entre a comunidade ajuda na cibertaxonomia dos dados, 97% das anotações feitas eram referentes ao nome científico (CRIA, 2019; CANHOS *et al.*, 2022).

Tendo em vista as limitações presentes dentro do SpeciesLink e a constante atualização de informações sobre os registros através das anotações, se faz necessário que se faça a verificação dos dados com os quais se está trabalhando. Além disso, é importante que as coleções que corroboram com dados padronizem a forma de acrescentar dados ao *database* para que seja possível diminuir a ambiguidade das informações e diminuir os possíveis erros presentes nos dados acrescentados.

Para além dos problemas relacionados a registros duplicados que não são lidos como duplicatas, há também grande ocorrência de problemas relacionados a classificação taxonômica dos registros, como nomes escritos de forma errada, nomes desatualizados, junção de famílias e gêneros distintos, nomenclaturas inexistentes e até mesmos imagens de plantas muito distintas representando a mesma entidade taxonômica.

Kennedy, Kukla & Paterson (2005) apontam a problemática ligada a identificação errada e/ou ambígua presente em sistemas de informações e bancos de dados online. A maior parte da imprecisão nos registros de espécies em bancos de dados são causados por nomenclaturas desatualizadas, o que resulta em diferentes combinações de nomes não válidos de uma espécie, tendo repetições por desatualização e presença de sinônimos taxonômicos no meio dos dados. Devido à grande inconsistência taxonômica presente nos registros em bancos de dados de biodiversidade, portanto, se faz necessário que se evite possíveis equívocos em relação a informação e a formação de uma perspectiva limitada sobre os dados, o que requer uma maior cautela ao se lidar com esses dados (FREITAS *et al.*, 2020).

4 CONCLUSÃO

Sistemas de informações como o SpeciesLink podem ser muito úteis para a pesquisa, educação e criação de políticas públicas, entretanto a presença de erros e ambiguidade nos dados obtidos pode levar a interpretações equivocadas sobre a informação e interpretações limitadas sobre os dados coletados.

Tendo em vista os problemas existentes quando se utiliza *databases* se faz necessário um processo de refinamento de dados criteriosos feito com cautela para que se possa obter dados mais seguros e menos incongruentes.

Ao verificar os dados aqui analisados, identificamos a ocorrência de um número considerável de registros para a flora da região, além de pelo menos treze espécies exclusivas do Cerrado em Boa Nova, além de outras 666 espécies que podem ocorrer tanto no Cerrado como em outros domínios fitogeográficos. Conhecer mais sobre essa área e reconhecê-la como parte importante das UC é essencial para a conservação da biodiversidade da região.

REFERÊNCIAS

++speciesLink network. Disponível em: <<https://specieslink.net/search/>>. Acesso em: 13 fev. 2023.

BARBOSA, M. M. D. C. *et al.* Species richness and similarity of the flora on four lajedos in Boa Nova, Bahia, Brazil. **Journal of mountain science**, v. 20, n. 6, p. 1526–1539, 2023.

BRANDÃO, G. de S. **Composição florística de trepadeiras do parque nacional de Boa Nova, bahia, brasil**. 2014. Dissertação (Mestrado em Genética, Biodiversidade e Conservação) - Universidade Estadual do Sudoeste da Bahia. Vitória da Conquista. 2014. 115 f. Disponível em: <http://www2.uesb.br/ppg/ppgbc2/wp->

content/uploads/2017/12/Disserta%C3%A7%C3%A3o-final_GirlandeBrandao-watermark.pdf. Acesso em: 17 mar. 2023.

BRASIL. Decreto de 11 de junho de 2010. Cria o Parque Nacional e o Refúgio de Vida Silvestre de Boa Nova, no Estado da Bahia, e dá outras providências. 2010.

CANHOS, D. A. L. *et al.* speciesLink: rich data and novel tools for digital assessments of biodiversity. **Biota neotropica**, v. 22. 2022.

CRIA. **Rede de informação colaborativa sobre a biodiversidade brasileira**. 2019. Disponível em: <<https://splink.cria.org.br/documentos/speciesLinkAbout052019.pdf>>.

Flora e Funga do Brasil. Disponível em: <<https://floradobrasil.jbrj.gov.br/reflora/listaBrasil/ConsultaPublicaUC/ConsultaPublicaUC.do>>. Acesso em: 25 dez. 2022.

FONSECA, W. O. *et al.* A família Rubiaceae no Parque Nacional de Boa Nova, Estado da Bahia, Brasil. **Hoehnea**, [s. l.], 19 mar. 2023. Disponível em: <https://www.scielo.br/j/hoehnea/a/zRvyp6DyBfKzwgJ9zxsxf/>. Acesso em: 18 mar. 2023.

FREITAS, T. M. S. *et al.* How reliable are species identifications in biodiversity big data? Evaluating the records of a neotropical fish family in online repositories. **Systematics and biodiversity**, v. 18, n. 2, p. 181–191, 2020.

International plant names index. Disponível em: <<https://www.ipni.org/>>. Acesso em: 3 mar. 2023.

KENNEDY, J. B.; KUKLA, R.; PATERSON, T. Scientific names are ambiguous as identifiers for biological taxa: Their context and definition are required for accurate data integration. Em: **Lecture Notes in Computer Science**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005. p. 80–95.

REGÔ, H. T. do; AZEVEDO, C. O. de. Sinopse das Orchidaceae do Parque Nacional de Boa Nova, BA, Brasil. **Hoehnea**, [s. l.], 19 mar. 2023. Disponível em: <https://www.scielo.br/j/hoehnea/a/3ny9LMzTztYwyQJJnmJQ5tL/abstract/?lang=pt>. Acesso em: 18 mar. 2023.

SECRETARIA DO MEIO AMBIENTE DA BAHIA; INSTITUTO DO MEIO AMBIENTE E RECURSOS HÍDRICOS (INEMA). Álbum seriado de educação ambiental: a sociobiodiversidade na Bahia. Salvador, 2017. Disponível em: <http://www.meioambiente.ba.gov.br/arquivos/File/EA/AlbumSociobiodiversidade.pdf>. Acesso em: 24/07/2023.

SIMÕES, S. dos S.; ZAPPI, D. C.; AONA, L. Y. S. A família Cactaceae no Parque Nacional de Boa Nova, Estado da Bahia, Brasil. **Hoehnea**, [s. l.], 19 mar. 2023. Disponível em: <https://www.scielo.br/j/hoehnea/a/754hs5stwgbGKPt5QNwhwFq/>. Acesso em: 18 mar. 2023.