

Checagem de nomes científicos no *speciesLink*

O *speciesLink* utiliza **dicionários** de nomes científicos para checar e atribuir status taxonômico aos nomes nas coleções da rede. Esses dicionários são atualizáveis e condizem com a natureza dinâmica da taxonomia.

Dicionários utilizados:

- **Flora e Funga do Brasil**

Versão mais recente disponível no IPT do JBRJ

https://ipt.jbrj.gov.br/jbrj/resource?r=lista_especies_flora_brasil

- **MycoBank**

Versão em tabela disponível na página principal do MycoBank

<https://www.mycobank.org/Images/MBList.zip>

- **AlgaeBase**

Versão de julho de 2019, sem atualizações recentes.

- **Catálogo de Abelhas Moure**

Última versão disponível (2022): <https://moure.cria.org.br>

- **LPSN** - List of Prokaryotic names with Standing in Nomenclature

Versão mais recente disponível em <https://lpsn.dsmz.de/> (necessário login).

- **GBIF Backbone**

Versão mais recente disponível em

<https://hosted-datasets.gbif.org/datasets/backbone/current/simple.txt.gz>

Atualização dos dicionários

Os dicionários são atualizados a cada seis meses, se houver novas versões. Após a atualização, todos os nomes científicos do *speciesLink* são novamente checados para garantir consistência. Não se faz atualização de dicionários individualmente.

Processo de checagem

1. **Escolha dos Dicionários:** A checagem de cada nome científico (família, gênero, espécie ou táxon infraespecífico) segue uma lógica para garantir que o status atribuído a ele seja único em toda a rede. Uma vez que um status é atribuído a um nome, essa informação é armazenada em *cache* e reutilizada sempre que aquele nome for encontrado na rede. Quando os dicionários são atualizados, o *cache* é apagado e o processo de checagem é reiniciado.

A sequência de dicionários utilizados na busca pelo nome científico é determinada, primeiramente, pelo Reino atribuído ao registro. Caso o Reino não exista ou não seja 'Fungi', 'Plantae' ou 'Animalia', utiliza-se o grupo em que a coleção foi classificada no *speciesLink*:

Fungi -> MycoBank, GBIF, Flora2020
Plantae -> AlgaeBase, MycoBank, Flora2020 e GBIF
Animalia -> Moure, GBIF

Caso o Reino não esteja definido ou seja outro, a classificação da coleção é utilizada para definir os dicionários:

Botânica -> AlgaeBase, MycoBank, Flora2020 e GBIF
Zoológica -> Moure, GBIF
Microbiológica -> LPSN, GBIF

Caso nenhum critério seja atendido, é usada a sequência completa:

AlgaeBase, MycoBank, Flora2020, Moure, LPSN, GBIF

Uma vez estabelecida a ordem em que os dicionários serão checados, o processo se inicia.

2. Preparação dos Nomes: Os nomes científicos e os autores são checados. Na rede *speciesLink*, os nomes científicos consistem de um (gênero), dois (binômio; gênero + epíteto específico) ou três termos (trinômio; gênero + epíteto específico + epíteto infra-específico). Exemplos de nomes que aparecem no *speciesLink*:

1. *Casearia sessiliflora* Cambess.
2. *Drimys brasiliensis* subsp. *sylvatica* (A.St.-Hil.) Ehrend. & Gottsb.
3. *Miconia albicans* (Sw.) Steub.
4. *Cassia desvauxii brevipes* Collad.
5. *Cassia javanica*
6. *Elaphoglossum* Schott ex J.Sm.

A busca pelos nomes desconsidera maiúsculas/minúsculas, bem como as designações dos epítetos infra-específicos. Os nomes dos autores também são reduzidos para evitar que espaços, maiúsculas/minúsculas, "e" ao invés de "&", e outros caracteres especiais restrinjam a busca. Assim, nossa lista final para efeito desta análise, torna-se assim:

1. *casearia sessiliflora cambess*
2. *drimys brasiliensis sylvatica asthilehrendgottsb*
3. *miconia albicans swsteub*
4. *cassia desvauxii brevipes collad*
5. *cassia javanica*
6. *elaphoglossum schottexjism*

3. Etapas da Checagem:

- Se o nome e autor já estiverem no cache, o status é atribuído.
- Se não, o sistema busca o nome no primeiro dicionário da lista. Se encontrado, o status é atribuído e o nome é armazenado no cache.
- Se não for encontrado, o processo é repetido em cada dicionário.
- Se ainda assim o nome não for encontrado, a busca continua sem considerar o autor.
- Se o **nome sem autor** não for encontrado em nenhum dicionário, ele é marcado como “não encontrado”.

Apresentação dos nomes no *speciesLink*

- Nomes encontrados com autor: **em negrito**.
- Nomes encontrados sem autor: sem negrito e com asterisco (*).
- Nomes aceitos: em **verde**.
- Sinônimos: em **cinza**.
- Nomes não encontrados: em **laranja**.
- Nomes com diferentes autores/status no mesmo dicionário (e.g., mesmo nome dado como aceito e sinônimo; homônimos): em **violeta** (considerados ambíguos).
- Registros identificados apenas até gênero: em **azul**.

Exemplos de exibição:

1. ***Casearia sessiliflora*** Cambess. — Aceito, com autor (Verde, negrito)
2. ***Drimys brasiliensis subsp. sylvatica*** (A.St.-Hil.) Ehrend. & Gottsb. — Aceito, com autor (Verde, negrito)
3. *Miconia albicans* (Sw.) Steub.* — Ambíguo (Violeta, sem negrito)
4. *Cassia desvauxii brevipes* Collad.* — Sinônimo (Cinza, sem negrito)
5. *Cassia javanica** — Aceito, sem autor (Verde, sem negrito)
6. *Elaphoglossum* Schott ex J.Sm. — Aceito, com autor (Azul, negrito)

Entendendo:

- No itens 1 e 2, o nome foi encontrado como nome aceito na Flora e Funga do Brasil, com o autor especificado pela coleção, por isso, **verde e negrito**.
- No item 3, o nome foi encontrado na Flora e Funga do Brasil mas com outros autores, sendo um aceito e outro sinônimo. Observe que o nome do autor tem a diferença de uma letra!
 - *Miconia albicans* (Sw.) Triana sinônimo de *Miconia albicans* (Sw.) Steud.
 - *Miconia albicans* (Sw.) Steud. como nome aceito
 Por isso, o nome é grafado no *speciesLink* em **violeta, sem negrito**.

- No item 4, o nome foi encontrado na Flora e Funga do Brasil como sinônimo, porém o nome do autor não foi encontrado. Por isso, o nome é grafado em cinza sem negrito e com asterisco (*).
- No item 5, o nome foi encontrado na Flora e Funga do Brasil como aceito, porém o nome do autor não foi informado pela coleção, portanto, não foi considerado na checagem. Por isso, o nome é grafado em verde, mas sem negrito e com asterisco (*).
- No item 6, o nome foi encontrado na Flora e Funga do Brasil, com o autor especificado pela coleção. Por ser um nome de gênero, aparece em azul.

Filtros de status taxonômico

O *speciesLink* oferece oito **filtros de status taxonômico** na sua interface de busca: **aceito com autor**, **aceito sem autor***, **sinônimo com autor**, **sinônimo sem autor***, **ambíguo com autor**, **ambíguo sem autor***, **não encontrado** e **só até gênero**.

Os usuários podem selecionar qualquer um desses filtros durante a busca, inclusive combinando múltiplos filtros, que podem ser utilizados juntamente com qualquer outro filtro do *speciesLink*. Por exemplo, se um pesquisador deseja gerar uma lista com apenas os nomes aceitos de plantas para o município de Campinas, ele deve: 1) preencher o campo de “**município**” com “Campinas” na seção “coleta”, 2) selecionar o “**tipo**” de coleção como “botânica” e, 3) marcar os filtros de status taxonômico “**aceito com autor**” e “**aceito sem autor**”. Caso queira filtrar apenas as espécies arbóreas dessa lista, basta selecionar a “**forma de vida**” como “Árvore” nos filtros chamados “flora & funga do Brasil”.

Data Cleaning: os filtros de status taxonômico também são valiosos para a limpeza e melhoria dos dados das coleções. Os curadores podem selecionar a coleção em que estão trabalhando através do filtro “**acrônimo**” em “**coleção**”. Em seguida, é possível escolher qualquer tipo de status de nome, sendo que os nomes encontrados sem autor (**aceitos***, **sinônimos*** ou **ambíguos***), nomes **ambíguos com autor**, **não encontrados** e **só até gênero** são especialmente promissores para data cleaning. Outra abordagem é selecionar apenas o acrônimo da coleção para filtrar todos os registros do acervo. Pode-se então optar por “**INVENTÁRIO**” para obter uma lista completa, que pode ser exportada em uma planilha. Se escolher “**REGISTROS**”, será gerada uma lista com todos os registros e o inventário ao lado. Em ambos os casos, é possível selecionar o “**status taxonômico**” no menu da tabela para filtrar as categorias e facilitar o processo de data cleaning. Para reduzir o número de registros, essa análise pode ser segmentada por famílias ou gêneros específicos, por exemplo.