

Porcentômetro

Sistema de Avaliação da Completude
Catálogo Taxonômico da Fauna do Brasil (CTFB)

Manual do Usuário — Português
Versão 2.1 - Abril de 2026

Acesse a aplicação em:

[https://comfortable-emotion-production-e366.up.railway.
app/](https://comfortable-emotion-production-e366.up.railway.app/)

O **Porcentômetro** é uma ferramenta web desenvolvida para o *Catálogo Taxonômico da Fauna do Brasil* (CTFB). Oferece dois tipos de análise complementares: (1) **Completude dos Dados** — grau de preenchimento dos campos obrigatórios das espécies catalogadas; e (2) **Completude da Lista de Espécies** — comparação dos registros do CTFB com o Global Biodiversity Information Facility (GBIF) para identificar espécies do Brasil ausentes no catálogo.

1. Acesso ao Sistema

Acesse a aplicação em qualquer navegador moderno:

```
https://comfortable-emotion-production-e366.up.railway.  
app/
```

Você verá a tela de login. Informe seu usuário e senha. Se este for seu primeiro acesso após ser cadastrado por um administrador, o sistema pedirá que você crie sua própria senha antes de continuar.

1.1 Autorização de Acesso

Cada usuário está associado a um ou mais táxons da planilha de pesquisadores (*pesquisadores.xlsx*). Você pode analisar qualquer táxon de sua área de responsabilidade e todos os grupos subordinados. Usuários vinculados a *Animalia* têm acesso a todo o catálogo. Administradores podem analisar qualquer táxon.

1.2 Primeiro Acesso — Criação de Senha

Usuários cadastrados pelo administrador via importação em lote **não possuem senha definida**. No primeiro acesso o sistema detecta isso e redireciona para a tela de criação de senha. Siga os passos:

1	Abra a URL e informe seu login. Deixe o campo de senha em branco (ou qualquer valor — será ignorado) e clique em <i>Entrar</i> .
2	O sistema redireciona para a tela Definir Senha com uma mensagem de boas-vindas.
3	Digite uma nova senha com no mínimo 6 caracteres , repita-a e clique em <i>Definir senha e entrar</i> .
4	Se as senhas coincidirem e atenderem ao tamanho mínimo, você é autenticado e redirecionado para a página principal.
5	Nos acessos seguintes, use seu login e a senha recém-criada.

Se as senhas não coincidirem ou forem curtas demais, uma mensagem de erro é exibida na mesma tela. Nenhum dado é perdido.

1.3 Senha Esquecida

Não é possível redefinir a senha esquecida por conta própria. Duas opções:

- **Contate o administrador** — ao lado da sua conta no Painel Admin, ele pode *Redefinir senha* (apaga a senha e ativa o flag "deve definir senha", de modo que no próximo acesso você segue o fluxo do primeiro acesso) ou *Definir senha* diretamente para uma nova senha que ele lhe informa, a qual você pode alterar depois (ver §1.5).
- **Escreva para o suporte** — envie e-mail para faunadobrasilctfb@gmail.com com seu nome de usuário. A equipe solicitará ao administrador que redefina seu acesso.

1.4 Página Principal

Após o login, a página principal exibe os cartões de seleção de análise. Clique no cartão desejado, informe o nome do táxon (ou clique em um dos botões de táxons autorizados para preenchimento automático) e pressione o botão de ação. O rodapé também oferece um link *Manual* (abre este manual no seu idioma) e um link *Alterar senha* (ver §1.5).

A tela de login e a página principal exibem a data da cópia do catálogo em uso (e a data do cadastro de pesquisadores). O catálogo é atualizado automaticamente a partir da fonte oficial da Fauna do Brasil (JBRJ) — ver §4.4.

1.5 Alterando Sua Senha

Após autenticado, você pode alterar sua própria senha a qualquer momento. Clique em *Alterar senha* no rodapé da página principal, informe a senha atual e, em seguida, a nova senha duas vezes (no mínimo 6 caracteres) e confirme. A alteração tem efeito imediato e passa a ser usada em todos os acessos seguintes. Funciona tanto para senhas que você mesmo criou quanto para uma senha definida por um administrador.

2. Análise de Completude dos Dados

Esta análise avalia a fração de campos obrigatórios preenchidos para cada espécie aceita (*taxonomicstatus* = *ACCEPTED*, *taxonrank* = *SPECIES* ou *SUBSPECIES*) dentro do táxon solicitado.

Regra de subespécies: quando uma espécie possui subespécies aceitas, a linha da espécie é excluída e apenas as subespécies são usadas como unidades de análise. Os resultados são agregados por família, ordem, classe, subclasse e filo. Quando o táxon de entrada é uma família, é gerada também uma agregação por gênero.

A análise é executada em segundo plano no servidor. Após enviar, uma página de status é exibida com um indicador de progresso e se atualiza automaticamente; ao concluir, redireciona para a página de resultados. Táxons grandes (ex: Hexapoda, Coleoptera) podem levar vários minutos — você pode deixar a aba aberta e a página avança sozinha. Se algo falhar, o erro é exibido na página de status, sem interromper a aplicação.

2.1 Os 15 Campos Avaliados

Os campos a seguir são avaliados para cada espécie (sujeitos às regras de exclusão ecológica da §2.2):

Campo	Descrição
<code>bibliographicreference</code>	Referência bibliográfica primária do táxon.
<code>country</code>	País de ocorrência (Brasil = BR).
<code>establishmentmeans</code>	Nativa, invasora, introduzida, etc.
<code>endemicbrazil</code>	Indica se a espécie é endêmica do Brasil.
<code>environment</code>	Tipo de ambiente: epicontinental, marinho ou ambos.
<code>locality</code>	Descrição da localidade de ocorrência.
<code>epicontinentaldomain</code>	Domínio fitogeográfico (espécies dulcícolas/terrestres).
<code>marinedomain</code>	Província biogeográfica marinha (espécies marinhas).
<code>lifeform</code>	Forma de vida / guilda alimentar.
<code>habitat</code>	Descrição do habitat.
<code>vernacularname</code>	Nome popular em português.
<code>vernacularnamelanguage</code>	Idioma do nome popular.
<code>vernacularnamelocality</code>	Região à qual o nome popular se refere.
<code>animalhostnames</code>	Espécies hospedeiras animais (ver §2.2 para regras de exclusão).
<code>vegetalhostnames</code>	Espécies hospedeiras vegetais (ver §2.2 para regras de exclusão).

2.2 Regras de Exclusão Ecológica

Determinados campos são excluídos do numerador e do denominador quando são biologicamente inaplicáveis. O denominador efetivo varia por espécie. A detecção de lifeform usa correspondência por substring para lidar com valores compostos (ex: `herbivoro;vida livre individual`).

Campos de domínio geográfico e país

Campo excluído	Condição	Justificativa
<code>marinedomain</code>	<code>environment = epicontinental</code>	Espécies dulcícolas/terrestres não possuem província biogeográfica marinha.
<code>epicontinentaldomain</code>	<code>environment = marinho</code>	Espécies estritamente marinhas não possuem domínio fitogeográfico continental.
um dos dois campos de domínio	<code>environment = vazio</code> ou em branco	Sempre exatamente UM slot de domínio por espécie. Para ambiente desconhecido, mantém o domínio que estiver preenchido; se ambos ou nenhum estiver preenchido, <code>marinedomain</code> é removido por padrão. Além disso, o próprio campo <code>environment</code> é tratado como não preenchido quando seu valor é "vazio".
<code>country</code>	<code>endemicbrazil = SIM</code>	Espécies endêmicas do Brasil sempre ocorrem no Brasil — o campo não acrescenta informação.

Campos de hospedeiro

O tratamento de `animalhostnames` e `vegetalhostnames` depende do valor de `lifeform` (correspondência por substring aplicada para lidar com valores compostos):

Valor de lifeform	Tratamento
contém <code>endoparasito</code> ou <code>ectoparasito</code>	Os dois campos são fundidos em um único slot de hospedeiro . O slot é preenchido se <i>qualquer um</i> não estiver vazio; vazio se ambos estiverem em branco. O denominador ganha 1 slot (não 2).
contém <code>herbivoro</code> (não-parasito)	<code>animalhostnames</code> é excluído (herbívoros se alimentam de plantas). <code>vegetalhostnames</code> é avaliado normalmente.
contém <code>vida livre individual</code> (não-parasito)	Ambos os campos de hospedeiro são excluídos — espécies de vida livre não possuem hospedeiros.
em branco / não preenchido	Ambos os campos de hospedeiro são excluídos — lifeform desconhecido, relação com hospedeiro indefinida.
qualquer outro valor (ex: <code>predador</code> , <code>colonial</code> , <code>sessil</code>)	Ambos <code>animalhostnames</code> e <code>vegetalhostnames</code> são incluídos normalmente no denominador.

Exemplos: (1) Um besouro de água doce (*environment = epicontinental*) é avaliado em 14 campos — *marinedomain* removido. (2) Um endoparasito é avaliado em 14 campos — os dois campos de hospedeiro são substituídos por um único slot combinado. (3) Um predador de vida livre (*lifeform = predador*) é avaliado em 15 campos — ambos os campos de hospedeiro contam normalmente. (4) Uma espécie com *lifeform = vida livre individual* é avaliada em 13 campos — ambos os campos de hospedeiro excluídos. (5) Um endêmico do Brasil (*endemicbrazil = SIM*) é avaliado em um campo a menos — *country* removido.

2.3 Fórmula de Completude

Para cada espécie / subespécie:

$$\text{completude (\%)} = (\text{campos aplicáveis preenchidos}) / (\text{total de campos aplicáveis}) \times 100$$

Os valores são agregados por grupo taxonômico como média, mínimo e máximo. **Amplitude** (máx – mín) mede a heterogeneidade interna do grupo.

2.4 Regra de Subespécies

Quando uma espécie possui pelo menos uma subespécie aceita (*taxonrank = SUBSPECIES*, *taxonomicstatus = ACCEPTED*), a linha da espécie é **excluída** da análise e suas subespécies aceitas são usadas como unidades de análise. Isso garante que a completude seja medida na resolução mais fina disponível para cada linhagem.

2.5 Atribuição de Pesquisador

Cada linha agregada lista o pesquisador (*nome_completo*) responsável pelo grupo. A busca começa no rank da própria linha e sobe na hierarquia até encontrar um responsável. Quando resolvido em rank superior, a célula é anotada com "(via <rank>)".

2.6 Arquivos de Saída

Os resultados são entregues em um arquivo ZIP contendo:

Arquivo	Conteúdo
completeness_per_species.xlsx	Uma linha por espécie / subespécie: <i>taxon_analisado</i> , contexto taxonômico, <i>taxonrank</i> , <i>completude (%)</i> , <i>filled_count</i> , <i>total_applicable</i> , pesquisador.
completeness_by_genus.xlsx	Agregação por gênero (apenas quando o táxon de entrada é uma família): <i>n_spp_ctfb</i> , <i>completude média/mín/máx</i> , <i>amplitude</i> , pesquisador.
completeness_by_family.xlsx	Agregação por família: mesmas colunas.
completeness_by_order.xlsx	Agregação por ordem.
completeness_by_class.xlsx	Agregação por classe.

completeness_by_subclass.xlsx	Agregação por subclasse (omitido se não houver dados).
completeness_by_phylum.xlsx	Agregação por filo (omitido em consultas de filo único).
completeness_histograms_genus.pdf	Gráfico de barras horizontais por gênero (apenas quando o táxon de entrada é uma família), com barras de erro mín–máx.
completeness_histograms_family.pdf	Gráfico por família. A barra de maior média é destacada em vermelho.
completeness_histograms_order.pdf	Mesmo gráfico por ordem.
completeness_histograms_class.pdf	Mesmo gráfico por classe.
completeness_histograms_subclass.pdf	Mesmo gráfico por subclasse.
completeness_histograms_phylum.pdf	Mesmo gráfico por filo.

3. Completude da Lista de Espécies (GBIF)

Esta análise compara a lista de espécies aceitas no CTFB com as espécies registradas no Brasil no GBIF, identificando espécies que ocorrem no Brasil mas estão ausentes do catálogo. A análise é executada em segundo plano; a página de status se atualiza automaticamente a cada 2 segundos.

3.1 Como Funciona — Quatro Fases

Fase 1 — Resolução do táxon

O nome do táxon é resolvido para um identificador interno do GBIF (*usageKey*) via a API GBIF Species Match (api.gbif.org/v1/species/match). Um erro é retornado se o nome não for reconhecido.

Fase 2 — Coleta de chaves de espécies

A API de Busca de Ocorrências do GBIF é consultada com *country=BR* e *taxonKey=<chave resolvida>*, usando facetas em *speciesKey* para recuperar todas as espécies com pelo menos um registro no Brasil. A paginação é aplicada automaticamente para táxons com mais de 50 000 espécies.

Fase 3 — Resolução de nomes

Cada chave de espécie é resolvida para um binômio canônico (gênero + epíteto específico) e sua taxonomia superior via a API de Espécies do GBIF. Até 20 requisições paralelas são feitas simultaneamente. Espécies fósseis/extintas (*extinct = true* no GBIF) são excluídas nesta fase.

Fase 4 — Comparação com o CTFB

As espécies do CTFB com *taxonomicstatus = ACCEPTED* e *taxonrank = SPECIES* são extraídas como binômios canônicos em minúsculas. Cada binômio do GBIF é classificado como: **presente** (corresponde a um nome aceito), **sinônimo** (corresponde a um sinônimo no CTFB) ou **ausente** (não encontrado de nenhuma forma).

3.2 Método de Correspondência e Ressalvas

A correspondência é feita pelo **binômio canônico em minúsculas**. Pontos importantes:

- Espécies com binômios idênticos no GBIF e no CTFB são contadas como presentes, independentemente do autor ou ano.
- Um nome do GBIF que corresponde a um sinônimo (registro não-ACCEPTED) no CTFB é classificado em *Sinonimos_no_CTFB*, não como ausente.
- Discrepâncias nomenclaturais — sinonímia de gênero, variações ortográficas, tratamentos taxonômicos distintos — farão uma espécie aparecer como "ausente" mesmo que esteja representada sob outro nome no CTFB. Recomenda-se revisão manual.
- Táxons infraespecíficos (subespécies, variedades) são excluídos de ambos os lados.
- Espécies fósseis são excluídas (*extinct = true* no GBIF).

3.3 Arquivo de Saída

Um único arquivo XLSX com quatro abas:

Aba	Conteúdo
-----	----------

Ausentes_no_CTFB	Espécies GBIF (Brasil, não-fósseis) ausentes do CTFB em qualquer forma. Colunas: phylum_gbif, classe_gbif, ordem_gbif, familia_gbif, nome_cientifico, gbif_species_key. Ordenado por filo → classe → ordem → família → nome.
Sinonimos_no_CTFB	Espécies GBIF cujo binômio corresponde a um sinônimo (não-ACCEPTED) no CTFB. Mesmas colunas de Ausentes_no_CTFB.
Resumo_por_Familia	Resumo por família: total de espécies GBIF, ausentes, sinônimos e presentes. Ordenado por número de ausentes (decrecente).
Presentes_no_CTFB	Espécies GBIF correspondidas a um registro ACCEPTED no CTFB. Útil para verificação de cobertura.

3.4 Tempo Estimado de Execução

Escala do táxon	Espécies GBIF aprox.	Tempo estimado
Ordem pequena (ex: Nematomorpha)	< 200	10–30 segundos
Ordem média (ex: Araneae)	~5 000	2–5 minutos
Ordem grande (ex: Coleoptera)	~100 000	15–30 minutos
Filo completo (ex: Arthropoda)	> 200 000	1–3 horas

Você pode fechar a aba e retornar mais tarde; o processo continua no servidor até a conclusão ou o próximo redeploy.

4. Administração

Contas de administrador têm acesso ao Painel Admin pelo link no rodapé da página principal.

4.1 Gerenciamento de Usuários

- **Importação em lote** — cole uma lista de logins da *pesquisadores.xlsx*. Cada usuário é criado com o flag "deve definir senha". Logins não encontrados na planilha são rejeitados.
- **Adição individual** — crie um único usuário com senha imediatamente utilizável.
- **Exportar CSV** — baixe um backup de todas as contas incluindo senhas hashadas. Faça commit deste arquivo (*users_backup.csv*) no repositório antes de cada redeploy para que os usuários sejam restaurados automaticamente na inicialização.
- **Restaurar CSV** — reimporte usuários de um CSV exportado anteriormente. Contas existentes são ignoradas; os hashes de senha são restaurados.
- **Redefinir senha** — apaga a senha de um usuário e ativa o flag "deve definir senha". No próximo acesso o usuário segue o fluxo do primeiro acesso.
- **Definir senha** — define diretamente uma senha específica e imediatamente utilizável para um usuário (sem fluxo de primeiro acesso). Útil para usuários que já concluíram o primeiro acesso. O usuário pode alterá-la depois (§1.5).

Seed automático na inicialização: a cada inicialização o sistema lê *users_backup.csv* da raiz do projeto e insere os usuários que ainda não existem no banco (contas e senhas existentes nunca são sobrescritas). Exporte e faça commit deste arquivo após adicionar ou alterar usuários para garantir que sobrevivam aos red deployments.

4.2 Persistência do Banco de Dados

Backend	Persistência	Como ativar
PostgreSQL	Permanente — sobrevive a todos os red deployments.	Defina DATABASE_URL com uma string postgresql:// (ex: pelo Railway).
SQLite em volume	Permanente — armazenado em um volume persistente montado.	Padrão: o arquivo do banco fica no diretório <i>data/</i> . Quando esse diretório é um volume montado (ex: volume Railway em <i>/app/data</i>), sobrevive aos red deployments. O Painel Admin confirma "SQLite em volume persistente".
SQLite (efêmero)	Efêmero — dados perdidos a cada redeploy.	Ocorre apenas se o diretório <i>data/</i> não estiver em um volume persistente. Um banner de aviso aparece no Painel Admin.

4.3 Fontes de Dados

Dois conjuntos de dados são carregados na inicialização, ambos a partir das APIs oficiais da Fauna do Brasil (JBRJ). O **catálogo** é baixado diretamente; o **cadastro de pesquisadores** é montado mesclando o relatório de contribuição ao vivo da JBRJ com uma planilha curada do Google Sheets (ver §4.4). Cada resultado é armazenado em cache no diretório *data/* após um carregamento bem-sucedido, e a cópia em cache é usada automaticamente se uma fonte estiver temporariamente

indisponível.

Arquivo	Conteúdo
planilha_extendida.csv	Catálogo CTFB estendido (Darwin Core + campos manuais). Baixado da API PostgREST da Fauna do Brasil (JBRJ) (fauna.jbrj.gov.br/rest/mv_planilha_extendida), paginado em páginas de 5 000 registros. Fonte de todos os cálculos de completude.
pesquisadores.xlsx	Cadastro de pesquisadores: login, nome_completo, táxons, institution, email. login / nome_completo / e-mail e os táxons atribuídos vêm do relatório de contribuição da JBRJ; o campo <i>institution</i> vem da planilha curada do Google Sheets. Controla autorização e atribuição de pesquisador.

4.4 Mantendo os Dados Atualizados

Tanto o catálogo quanto o cadastro de pesquisadores são servidos a partir de cópias recentes das fontes oficiais da JBRJ, mantidas atualizadas automaticamente — nenhum upload manual é necessário.

- **A cada inicialização e, novamente, diariamente às 03:00 (America/São_Paulo)**, uma tarefa agendada rebaixa o catálogo e o relatório de contribuição de pesquisadores e recarrega ambos na memória, de modo que uma instância em execução prolongada nunca fique desatualizada. (O agendador pode ser desativado com a variável de ambiente `ENABLE_SCHEDULER=0`.)
- **Catálogo** — baixado integralmente da API da Fauna do Brasil. O download é atômico e à prova de falhas: o novo arquivo substitui o anterior apenas após um download completo e bem-sucedido; se a API estiver indisponível ou não retornar registros, a última cópia válida em cache é mantida e as análises continuam sem interrupção.
- **Cadastro de pesquisadores** — login, nome_completo, e-mail e os táxons atribuídos a cada pesquisador vêm do relatório de contribuição da JBRJ (`mv_relatorio_contribuicao_usuarios`); o campo *institution* e qualquer pesquisador ausente do relatório são mesclados a partir da planilha curada do Google Sheets. Se um pesquisador não tiver táxons no relatório, os táxons da planilha são mantidos como fallback.

As datas das cópias do catálogo e do cadastro de pesquisadores em uso são exibidas a todos os usuários tanto na tela de login quanto na página principal (ex: Catálogo: 2026-06-22 · Pesquisadores: 2026-06-22), para que você sempre confirme o quão atualizados estão os dados.

5. Interpretando os Resultados

5.1 Gráficos de Completude

Cada PDF exibe uma barra horizontal por grupo taxonômico. O comprimento da barra representa a completude média (%). As barras de erro se estendem até os valores mínimo e máximo do grupo. A barra de maior média é colorida em vermelho; as demais são azuis. Os valores são anotados à direita.

Um grupo com 100% de média e amplitude zero está perfeitamente homogêneo. Uma amplitude grande indica que algumas espécies são bem documentadas enquanto outras carecem de muitos campos — sinalizando onde concentrar o esforço curatorial.

5.2 Lista de Lacunas de Espécies

A aba *Ausentes_no_CTFB* lista espécies candidatas à inclusão no catálogo. Antes de agir sobre um registro, considere:

- Diferenças nomenclaturais — a espécie pode já estar no CTFB sob um binômio sinônimo (verifique primeiro a aba *Sinonimos_no_CTFB*).
- Escopo taxonômico — alguns registros do GBIF podem referir-se a espécies marinhas fora do escopo atual do CTFB.
- Qualidade dos registros — alguns registros do GBIF são identificações incorretas ou carecem de precisão geográfica. Registros de alta incerteza devem ser verificados.
- Táxons infraespecíficos — a comparação com o GBIF abrange apenas o nível de espécie.