

Name matching in GBIF

Cecilie Svenningsen | csvenningsen@gbif.org



Names on GBIF.org | searching names

The screenshot shows the GBIF.org search results page. The browser address bar is highlighted with a green box, showing the URL <https://www.gbif.org/species/search?q=>. The page header includes navigation links: Get data, How-to, Tools, Community, and About. A search bar and a Login button are also present. The main content area displays search results for species names, with a total of 7,651,854 results. The results are categorized by taxonomic rank, with 'BACKBONE' and 'ALL CHECKLISTS' tabs visible. The first three results are Kingdoms: Animalia (1,834,450,359 occurrences), Archaea (404,117 occurrences), and Bacteria (22,066,369 occurrences). Each result includes a small heatmap visualization. A green box highlights the 'BACKBONE' and 'ALL CHECKLISTS' tabs. Another green box highlights a message in the left sidebar: 'You are searching for species in the GBIF taxonomy. SEARCH ACROSS CHECKLISTS'.

Species

SEARCH SPECIES | 7,651,854 RESULTS

BACKBONE ALL CHECKLISTS

Animalia Kingdom

Accepted Kingdom 1,834,450,359 occurrences

Archaea Kingdom

Accepted Kingdom 404,117 occurrences

Bacteria Kingdom

Accepted Kingdom 22,066,369 occurrences

You are searching for species in the GBIF taxonomy.
SEARCH ACROSS CHECKLISTS

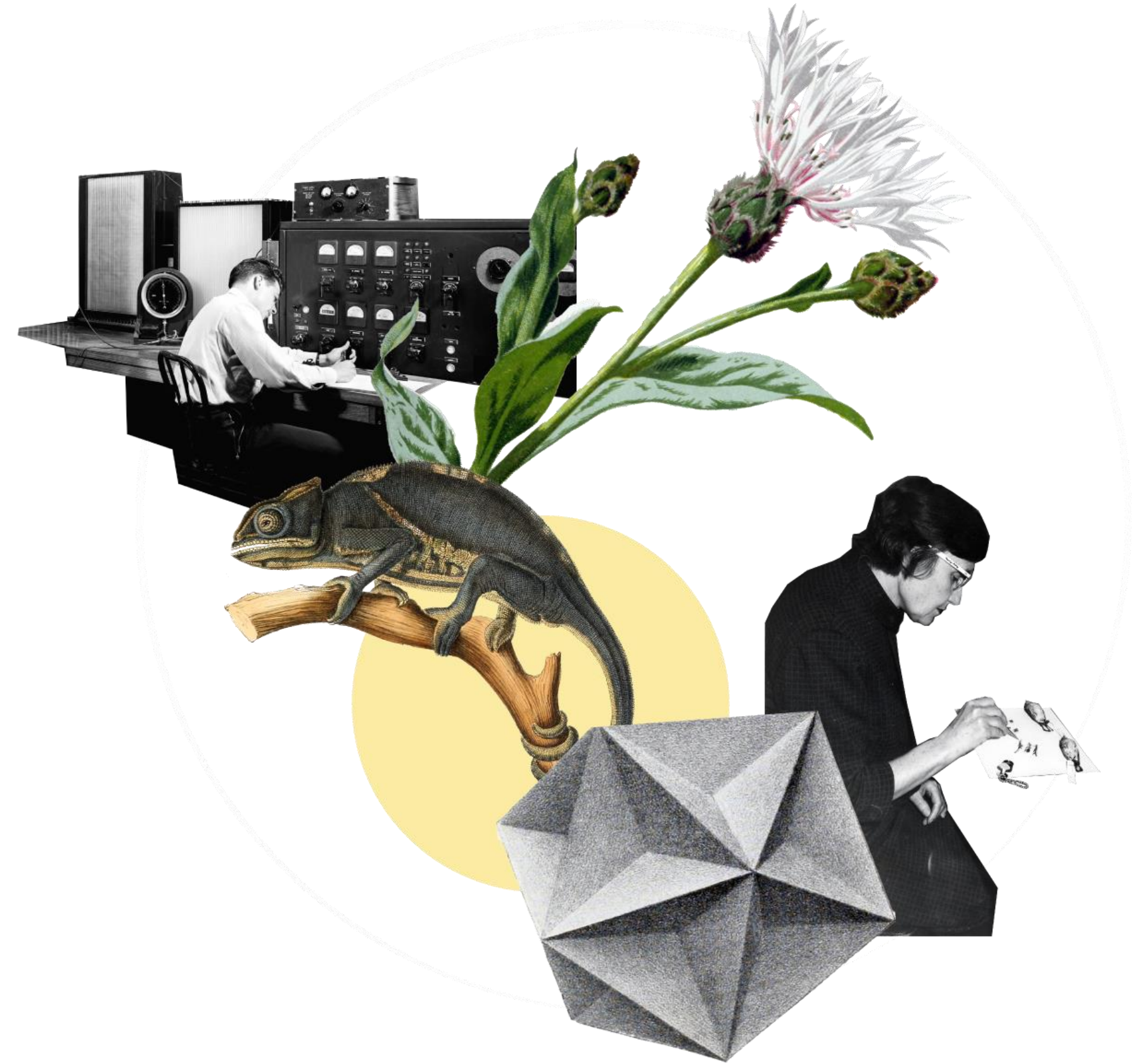
<https://www.gbif.org/species/search?q=>



Searching names on GBIF.org

The search function is a full text search of name usages covering

- the scientific and vernacular name
- the species description
- distribution
- the entire classification across all name usages of all or some checklists



Search API

query

mallard duck

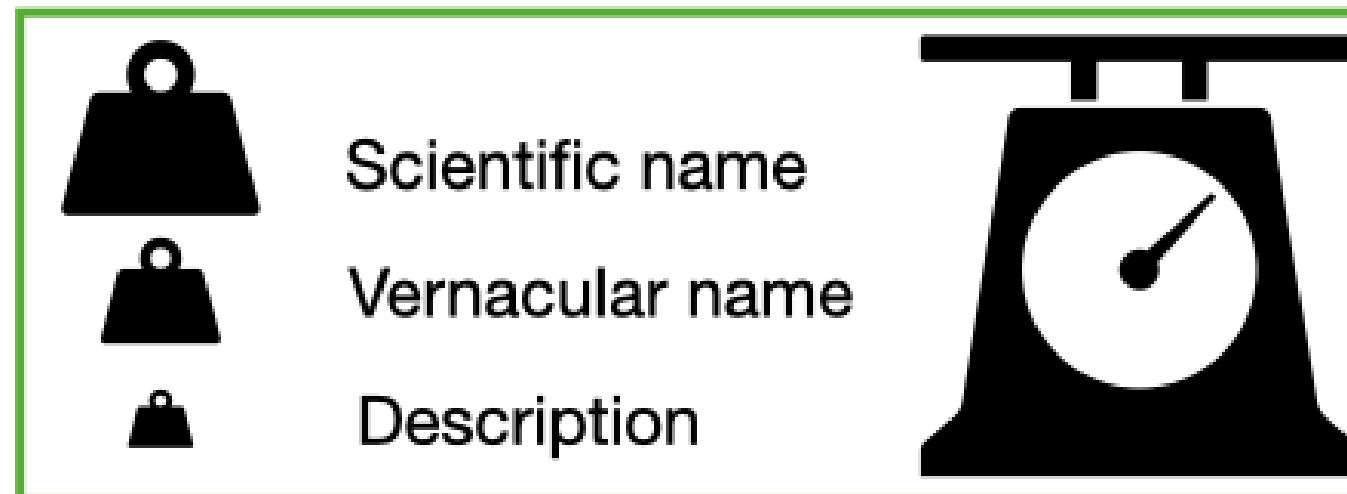
Queries with Lucene fuzzy search system

Indexed taxa texts

Scientific name	Description	Vernacular	...
Anas boschas	... in Europae...	-	
Morishitium feizullaevi	... of mallard duck ...	-	
Cyclocoelum oculeum	... hosts. Mallard duck ...	-	
Anas platyrhynchos	... CUDDLINESS cuddly...	mallard duck	
...	

Ranks Lucene matches

Ranking algorithm



Returns ranked matches

1. Anas platyrhynchos
2. Morishitium feizullaevi
3. Cyclocoelum oculeum



Browser tabs: Inbox, To-do, Dataset, Nature, Dataset, Manag, Dataset, Dataset, Nature, Species, (Almor, 17 Me, API, ppt-ter, ppt-ter, AsiaDL, GBIF, Occur, Occur, Spe X

Address bar: <https://www.gbif.org/species/search?q=>

160%

Get data How-to Tools Community About

Species

SEARCH SPECIES | 7,651,854 RESULTS

BACKBONE ALL CHECKLISTS

Search

Rank

Status

Higher taxon

Issues and flags

You are searching for species in the GBIF taxonomy.

SEARCH ACROSS CHECKLISTS

Animalia Kingdom
Accepted Kingdom 1,834,450,355 occurrences

Archaea Kingdom
Accepted Kingdom 404,117 occurrences

Bacteria Kingdom
Accepted Kingdom 22,066,369 occurrences

Type here to search

09:03 17-05-2023



Names on GBIF.org | searching occurrences

The screenshot displays the GBIF.org search interface. The browser address bar shows the URL <https://www.gbif.org/occurrence/search>. The page header includes navigation links: Get data, How-to, Tools, Community, and About. A search bar is located in the top right, and a 'Login' button is visible. The main content area is titled 'SEARCH OCCURRENCES | 2,347,469,935 RESULTS'. Below this, there are tabs for TABLE, GALLERY, MAP, TAXONOMY, METRICS, and DOWNLOAD. The 'TABLE' tab is selected, showing a table of search results. The table has columns for Scientific name, Country or area, Coordinates, Month & year, and Occurrence status. The results list various species such as *Sitta pusilla*, *Calidris alpina*, *Larus argentatus*, *Passer domesticus*, *Mergus merganser*, *Cyanocitta cristata*, *Rytidosperma racemosum*, *Lerista apoda*, and *Parus major*. On the left side, there is a sidebar with a search bar, filter options (Simple, Advanced), and a 'Major groups' section listing taxonomic categories and their counts.

Scientific name	Country or area	Coordinates	Month & year	Occurrence status
<i>Sitta pusilla</i> Latham, 1790	United States of America	32.3N, 81.8W	2023 January	Present
<i>Calidris alpina</i> (Linnaeus, 1758)	United Kingdom of Great ...	51.6N, 4.2W	2023 January	Present
<i>Larus argentatus</i> Pontoppidan, 1763	United Kingdom of Great ...	54.3N, 0.4W	2023 January	Present
<i>Passer domesticus</i> (Linnaeus, 1758)	United Kingdom of Great ...	54.5N, 1.9W	2023 January	Present
<i>Mergus merganser</i> Linnaeus, 1758	United Kingdom of Great ...	55.9N, 4.3W	2023 January	Present
<i>Cyanocitta cristata</i> (Linnaeus, 1758)	United States of America	29.5N, 95.1W	2023 January	Present
<i>Rytidosperma racemosum</i> (R.Br.) Connor & ...	Australia	35.2S, 149.0E	2023 January	Present
<i>Lerista apoda</i> Storr, 1976	Australia	17.9S, 122.3E	2023 January	Present
<i>Parus major</i> Linnaeus, 1758	United Kingdom of Great ...	52.2N, 0.8E	2023 January	Present

<https://www.gbif.org/occurrence/search>



Matching names on GBIF.org

The match function will do a

- fuzzy matching on canonical names
- generates candidate matches
- the scientific authorships are excluded for the first step of the matching
- the resulting matches are then scored according to their taxonomy, authorship, status and rank.



Match API

query

Name: Agathis montana
Kingdom: Plantae

Queries with Lucene fuzzy search system

Compute taxonomy, authorship, status and rank scores

Returns main match and alternatives or no match if the scores are the same

Normalized and indexed canonical Scientific Names

Canonical scientific Name	Authorship	Kingdom	...
Anas platyrhynchos	Linnaeus, 1758	Animalia	
Agathis montana	Shestakov, 1932	Animalia	
Agathis montana	de Laub.	Plantae	
Agathon montanus	(Kitakami, 1931)	Animalia	
Agallis montana	Phil.	Plantae	
...	

Normalized and indexed taxonomy, Authorship, rank, etc.

Canonical scientific Name	Authorship	Kingdom	...
Agathis montana	Shestakov, 1932	Animalia	
Agathis montana	de Laub.	Plantae	
Agathon montanus	(Kitakami, 1931)	Animalia	
Agallis montana	Phil.	Plantae	

Main match: Agathis montana de Laub.

Alternatives:

1. Agallis montana Phil.
2. Agathis montana Shestakov, 1932
3. Agathon montanus (Kitakami, 1931)



GBIF | Global Biodiversity Information Facility

Free and open access to biodiversity data

OCCURRENCES SPECIES DATASETS PUBLISHERS RESOURCES

What is GBIF? About GBIF Denmark

10:30 17-05-2023



The match function outputs data quality flags

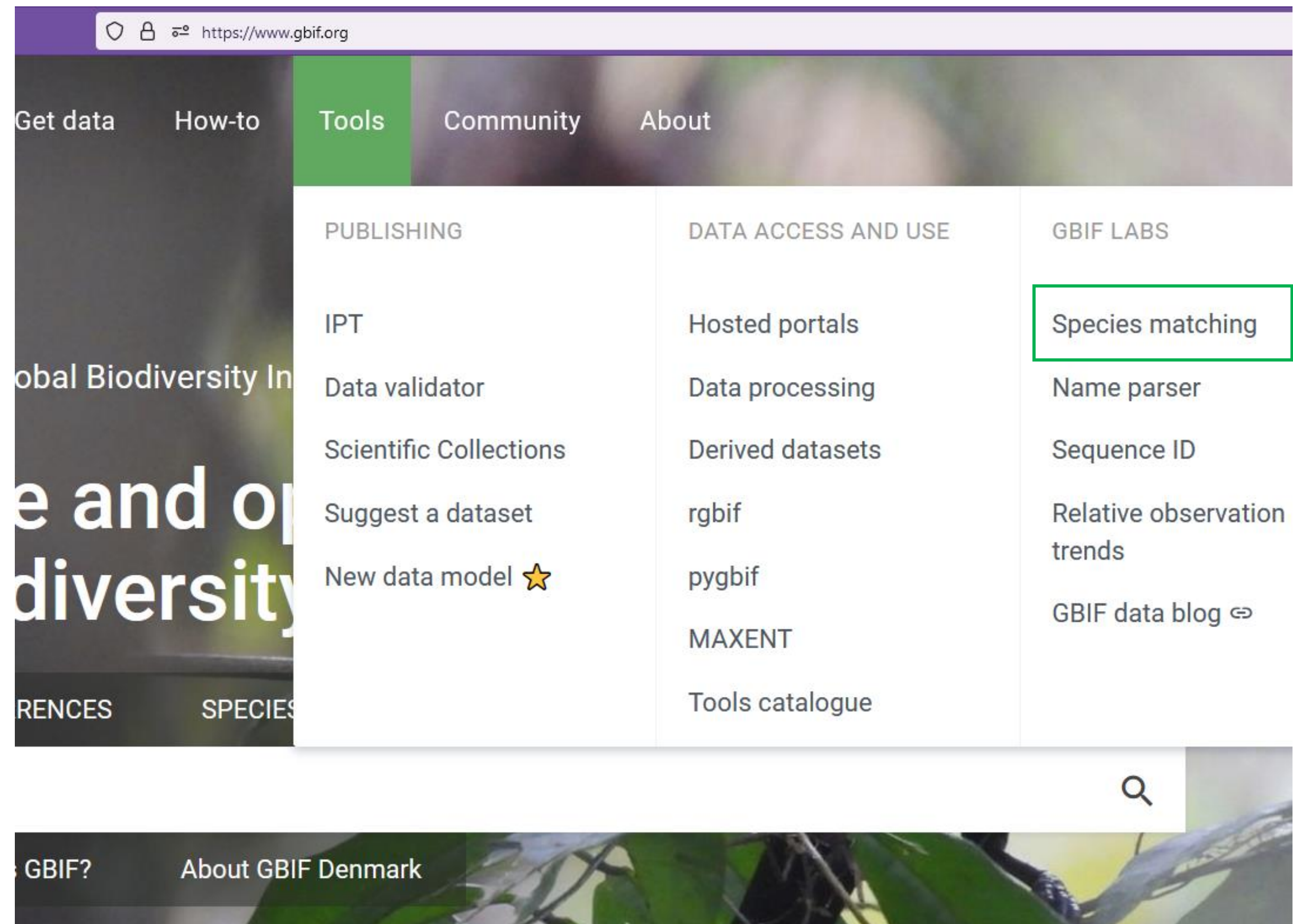
Data quality flags are associated with occurrences and as outputs of downloads or API calls. Relevant flags for names are:

- Taxon match fuzzy
- Taxon match higher rank
- Taxon match none



THE SPECIES MATCHING TOOL

A wrapper around the Species Match API





Get data

How-to

Tools

Community

About



Login

TOOLS | LOOK UP

Normalize species names from a csv file against the GBIF backbone.
The file is expected to be have a column called 'scientificName' and an optional column 'kingdom' and 'id'.

SIMPLEEXAMPLE.CSV

ADVANCEDEXAMPLE.CSV

SELECT FILE

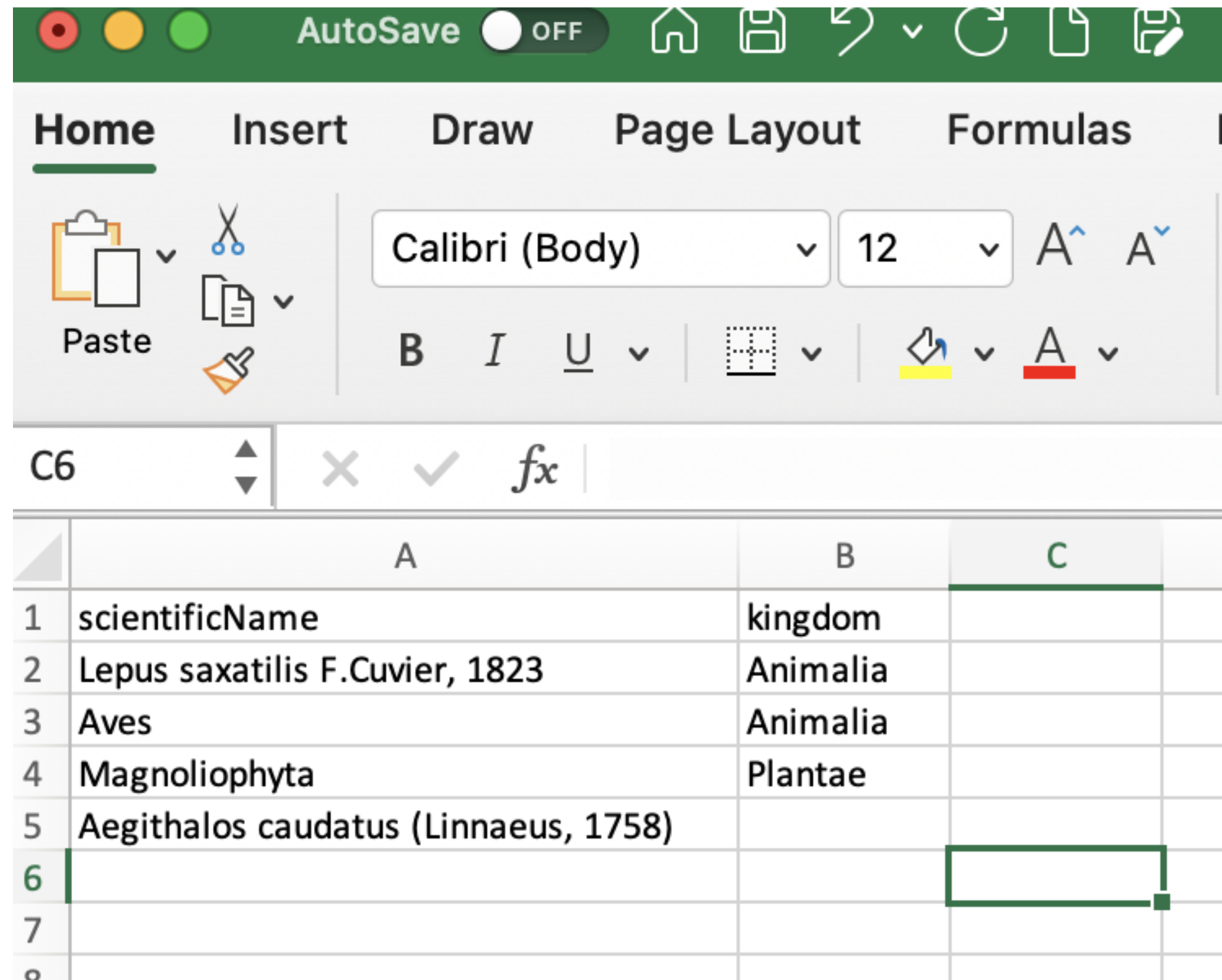
or

DROP HERE

<https://www.gbif.org/tools/species-lookup>



THE SPECIES MATCHING TOOL | formatting the input file



AutoSave OFF

Home Insert Draw Page Layout Formulas

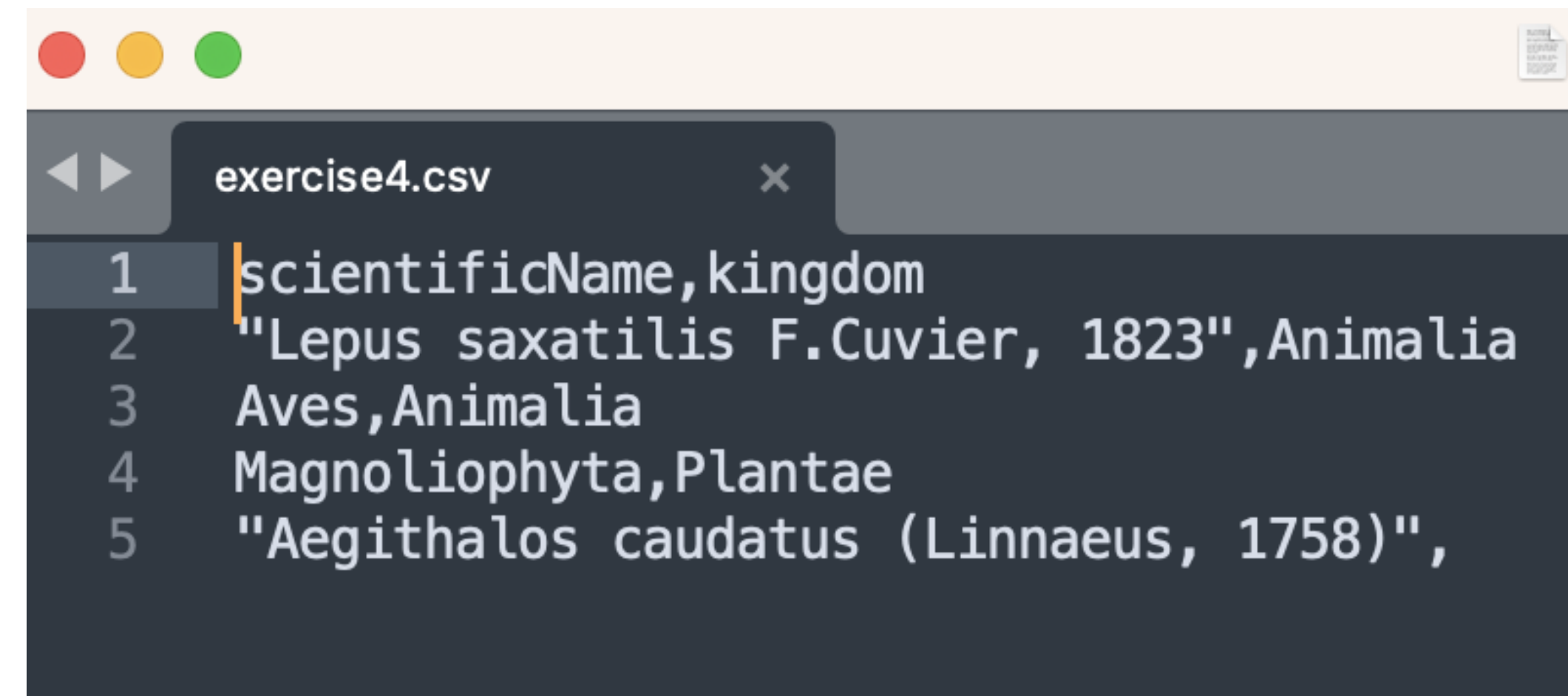
Paste

Calibri (Body) 12

B *I* U

C6

	A	B	C
1	scientificName	kingdom	
2	Lepus saxatilis F.Cuvier, 1823	Animalia	
3	Aves	Animalia	
4	Magnoliophyta	Plantae	
5	Aegithalos caudatus (Linnaeus, 1758)		
6			
7			
8			



```
exercise4.csv
1 scientificName,kingdom
2 "Lepus saxatilis F.Cuvier, 1823",Animalia
3 Aves,Animalia
4 Magnoliophyta,Plantae
5 "Aegithalos caudatus (Linnaeus, 1758)",
```



THE SPECIES MATCHING TOOL | interface after upload

Get data How-to Tools Community About Login

TOOLS | LOOK UP

If no kingdom specified then prefer

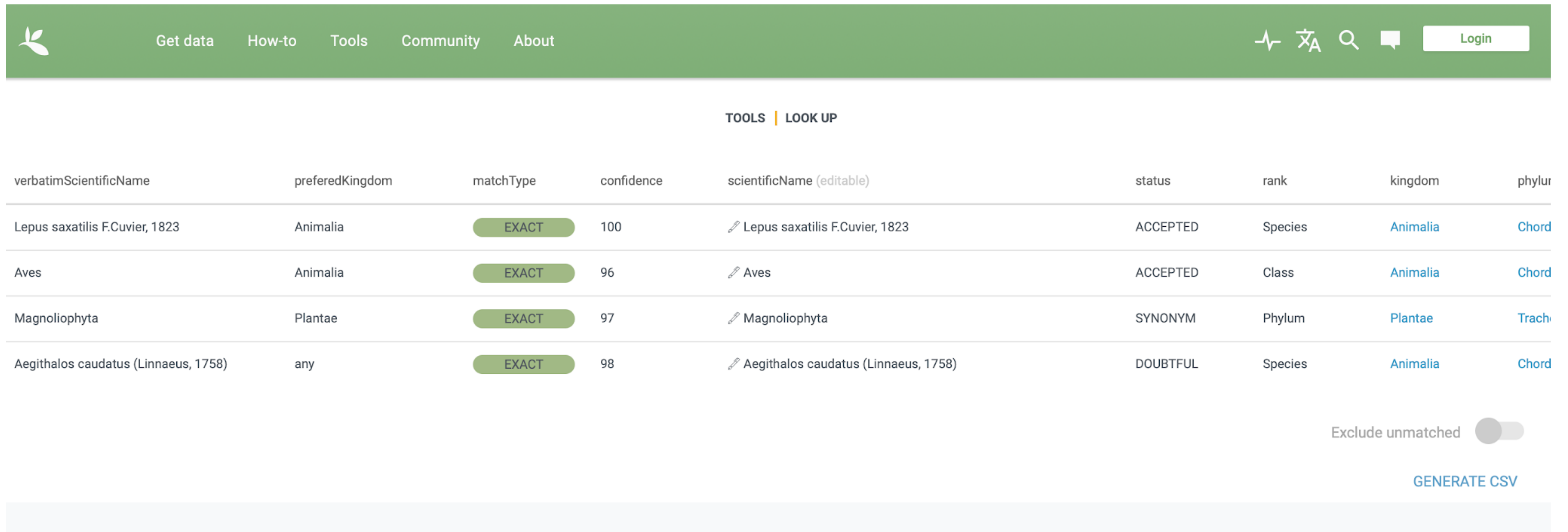
animalia plantae fungi chromista bacteria viruses archaea

MATCH TO GBIF BACKBONE

scientificName	preferred kingdom
Lepus saxatilis F.Cuvier, 1823	Animalia
Aves	Animalia
Magnoliophyta	Plantae
Aegithalos caudatus (Linnaeus, 1758)	







THE SPECIES MATCHING TOOL | interface after matching



The image shows the interface of the Species Matching Tool after a search. At the top, there is a green navigation bar with a logo on the left and links for 'Get data', 'How-to', 'Tools', 'Community', and 'About'. On the right side of the bar are icons for a home page, a search bar, a magnifying glass, a speech bubble, and a 'Login' button.

Below the navigation bar, the page is titled 'TOOLS | LOOK UP'. The main content is a table with the following columns: 'verbatimScientificName', 'preferredKingdom', 'matchType', 'confidence', 'scientificName (editable)', 'status', 'rank', 'kingdom', and 'phylur'. The table contains four rows of data:

verbatimScientificName	preferredKingdom	matchType	confidence	scientificName (editable)	status	rank	kingdom	phylur
Lepus saxatilis F.Cuvier, 1823	Animalia	EXACT	100	 Lepus saxatilis F.Cuvier, 1823	ACCEPTED	Species	Animalia	Chord
Aves	Animalia	EXACT	96	 Aves	ACCEPTED	Class	Animalia	Chord
Magnoliophyta	Plantae	EXACT	97	 Magnoliophyta	SYNONYM	Phylum	Plantae	Trach
Aegithalos caudatus (Linnaeus, 1758)	any	EXACT	98	 Aegithalos caudatus (Linnaeus, 1758)	DOUBTFUL	Species	Animalia	Chord

At the bottom right of the table area, there is a toggle switch labeled 'Exclude unmatched' which is currently turned off. Below the toggle is a blue button labeled 'GENERATE CSV'.



THE SPECIES MATCHING TOOL | example editing

The screenshot displays the Species Matching Tool interface. At the top, there is a green navigation bar with a logo on the left and links for 'Get data', 'How-to', 'Tools', 'Community', and 'About' in the center. On the right side of the navigation bar are icons for a heart, a magnifying glass, a search icon, and a 'Login' button.

The main content area is a table with columns: 'verbatimScientificName', 'preferredKingdom', 'match', 'status', 'rank', 'kingdom', and 'phylum'. The table contains several rows of data, including 'Lepus saxatilis F.Cuvier, 1823', 'Aves', 'Magnoliophyta', and 'Aegithalos caudatus (Linnaeus, 1758)'. The 'Aegithalos caudatus' row is highlighted, and a modal dialog is open over it.

The modal dialog is titled 'Magnoliophyta' and contains a search input field with the text 'Tracheophy|'. Below the input field, there is a dropdown menu showing 'Tracheophyta Phylum (Accepted)' and 'Plantae'. At the bottom of the modal, there are two buttons: 'CANCEL' and 'DISCARD'.

At the bottom right of the main interface, there is a toggle switch for 'Exclude unmatched' and a 'GENERATE CSV' button.

verbatimScientificName	preferredKingdom	match	status	rank	kingdom	phylum
Lepus saxatilis F.Cuvier, 1823	Animalia	EXACT	ACCEPTED	Species	Animalia	Chordata
Aves	Animalia	EXACT	ACCEPTED	Class	Animalia	Chordata
Magnoliophyta	Plantae	EXACT	SYNONYM	Phylum	Plantae	Tracheophyta
Aegithalos caudatus (Linnaeus, 1758)	any	EXACT 98	DOUBTFUL	Species	Animalia	Chordata



THE SPECIES MATCHING TOOL | download result

Get data How-to Tools Community About 🏠 🌐 🔍 🗨️ Login

TOOLS | LOOK UP

verbatimScientificName	preferedKingdom	matchType	confidence	scientificName (editable)	status	rank	kingdom	phylur
Lepus saxatilis F.Cuvier, 1823	Animalia	EXACT	100	Lepus saxatilis F.Cuvier, 1823	ACCEPTED	Species	Animalia	Chord
Aves	Animalia	EXACT	96	Aves	ACCEPTED	Class	Animalia	Chord
Magnoliophyta	Plantae	EDITED		Tracheophyta	ACCEPTED	Phylum	Plantae	Trachu
Aegithalos caudatus (Linnaeus, 1758)	any	EXACT	98	Aegithalos caudatus (Linnaeus, 1758)	DOUBTFUL	Species	Animalia	Chord

Exclude unmatched

GENERATE CSV



<https://www.gbif.org/tools/species-lookup>



THE SPECIES MATCHING TOOL | results

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	occurrenceId	verbatimScientificName	scientificName	key	matchType	confidence	status	rank	kingdom	phylum	class	order	family	genus	species
2		Lepus saxatilis F.Cuvier, 1823	Lepus saxatilis F.Cuvier, 1823	2436775	EXACT	100	ACCEPTED	SPECIES	Animalia	Chordata	Mammalia	Lagomorpha	Leporidae	Lepus	Lepus saxa
3		Aves	Aves	212	EXACT	96	ACCEPTED	CLASS	Animalia	Chordata	Aves				
4		Magnoliophyta	Tracheophyta	7707728	EDITED		ACCEPTED	PHYLUM	Plantae	Tracheophyta					
5		Aegithalos caudatus (Linnaeus, 1758)	Aegithalos caudatus (Linnaeus, 1758)	2495000	EXACT	98	DOUBTFUL	SPECIES	Animalia	Chordata	Aves	Passeriformes	Aegithalidae	Aegithalos	Aegithalos
6															
7															

