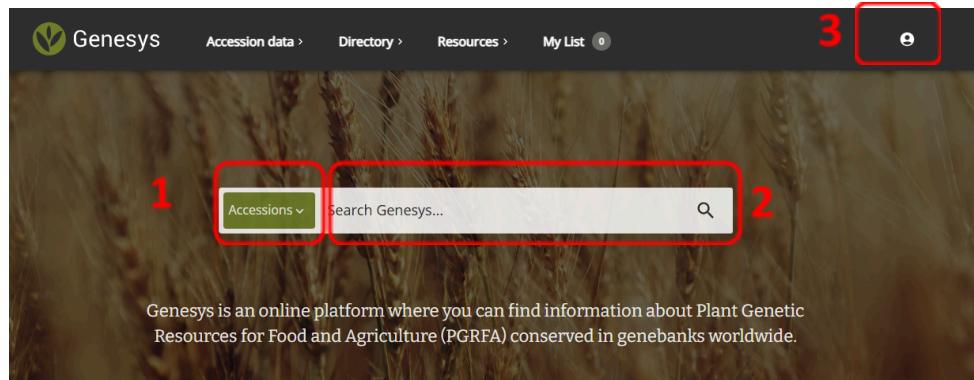


## DOWNLOAD OCCURRENCE DATA

This section illustrates the process for downloading occurrence data from the available data sources for a specific crop or taxon.

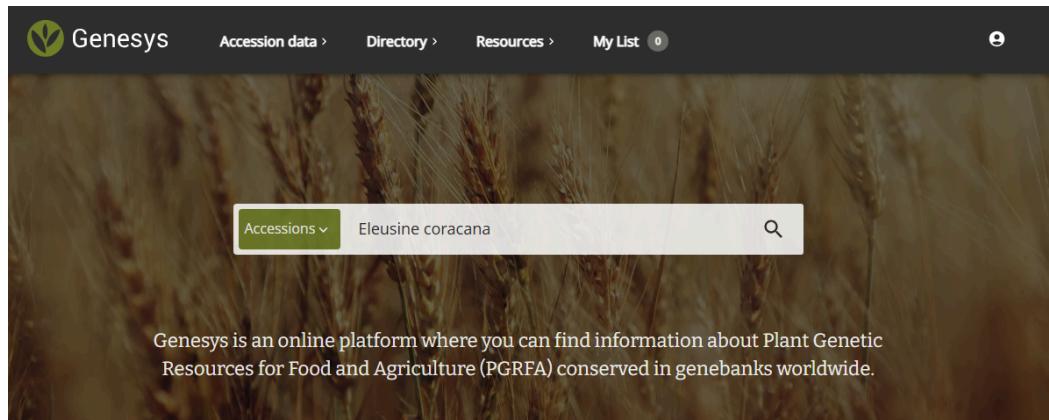
### DOWNLOAD OCCURRENCES FROM GENESYS-PGR (<https://www.genesys-pgr.org>):

Genesys-PGR (Plant Genetic Resources) is an online platform that provides access to information about Plant Genetic Resources for Food and Agriculture (PGRFA) conserved in gene banks worldwide. Launched in 2011, Genesys aims to facilitate the discovery, access, and use of ex situ collections of plant germplasm. It serves researchers, plant breeders, and other users by offering a unified gateway to explore the diversity of crops stored in gene banks. Genesys integrates data from various global networks and gene banks, detailing millions of accessions, enabling efficient searches, and supporting the conservation and sustainable use of crop diversity.



1. Type of data for query: Make sure “Accessions” is selected.
2. Search bar: field to type a query, user can type any information related to accessions (accessionID, genebankID, genus, species, etc.).
3. Login button: it is recommended to create an account and log in (Google authentication is also available).

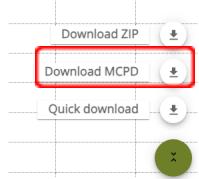
In this example we will query for finger millet (*Eleusine coracana*) accession data, we strongly recommend using the scientific name as query, then click on the magnified glass icon.



Once the genesys-pgr portal finish loading, we should see the next page.

	My List	Institute code	Holding Instit...	Accession number	Accession name	Taxonomy	Crop name	Crop	Biologic
1	<input type="checkbox"/>	KEN214	Seed Savers ...	SSN1000		<i>Eleusine coracana</i>			
2	<input type="checkbox"/>	KEN214	Seed Savers ...	SSN0127		<i>Eleusine coracana</i>			
3	<input type="checkbox"/>	KEN214	Seed Savers ...	SSN0112		<i>Eleusine coracana</i>			
4	<input type="checkbox"/>	KEN214	Seed Savers ...	SSN0134		<i>Eleusine coracana</i>			
5	<input type="checkbox"/>	KEN214	Seed Savers ...	SSN0140		<i>Eleusine coracana</i>			
6	<input type="checkbox"/>	KEN214	Seed Savers ...	SSN0142		<i>Eleusine coracana</i>			
7	<input type="checkbox"/>	KEN214	Seed Savers ...	SSN0145		<i>Eleusine coracana</i>			
8	<input type="checkbox"/>	KEN214	Seed Savers ...	SSN0118		<i>Eleusine coracana</i>			
9	<input type="checkbox"/>	KEN214	Seed Savers ...	SSN0912		<i>Eleusine coracana</i>			
10	<input type="checkbox"/>	KEN214	Seed Savers ...	SSN0919		<i>Eleusine coracana</i>			
11	<input type="checkbox"/>	KEN214	Seed Savers ...	SSN0865		<i>Eleusine coracana</i>			

1. Make sure "Manual filtering" tab is selected.
2. Filter bar: shows the available filters – after a filter is selected/modified make sure of click on "APPLY FILTERS" button to apply the filters.
3. Section with information of the filters that are currently applied and the number of records retrieved.
4. Passport data database visualization.
5. Download Button: Currently data can be downloaded using three different options. We recommend downloading the MCPD as illustrated below.



To refine the query further, we will select only accessions of *Eleusine coracana*. As shown, the data is currently filtered to include only the *Eleusine* genus.

Species	Count
<i>Eleusine coracana</i>	14,822
<i>Eleusine indica</i>	242
<i>Eleusine africana</i>	139
<i>Eleusine sp.</i>	70
<i>Eleusine tristachya</i>	39
<i>Eleusine jaegeri</i>	22
<i>Eleusine multiflora</i>	17

In the **Taxonomy** filter, we can search for the specific species to select (can select multiple species), in this case, we will choose “*Eleusine coracana*” from the **Species** filter

Species	Count
<i>Eleusine coracana</i>	14,822
<i>Eleusine indica</i>	242
<i>Eleusine africana</i>	139
<i>Eleusine sp.</i>	70
<i>Eleusine tristachya</i>	39
<i>Eleusine jaegeri</i>	22
<i>Eleusine multiflora</i>	17

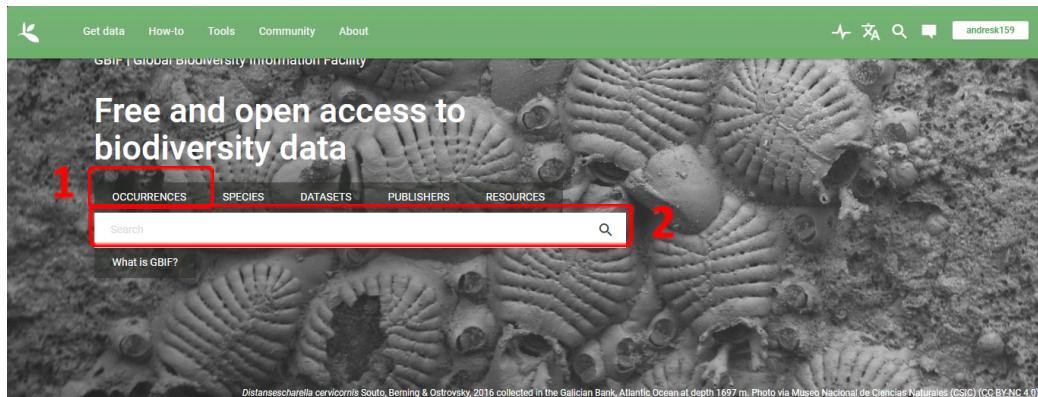
Once it's selected, click on “APPLY FILTERS” button to apply.

The screenshot shows the Genesys Accession browser interface. On the left, there's a sidebar with a 'TAXONOMY' section containing filters for 'Genus' (Eleusine), 'Species' (Eleusine coracana), and 'Specific epithet' (coracana). The main area is titled 'Accession browser' and displays a table of approximately 14,822 accessions. The columns include 'My List', 'Institute code', 'Holding instit...', 'Accession number', 'Accession name', 'Taxonomy', 'Crop name', 'Crop', and 'Biologic'. The first few rows show entries like KEN214, Seed Savers ..., SSN1000, Eleusine coracana.

Now, the results are restricted to 14.822 accessions. Now we can download the database.

## DOWNLOAD OCCURRENCES FROM GBIF(<https://www.gbif.org>):

GBIF stands for the Global Biodiversity Information Facility. It is an international network and data infrastructure funded by governments and organizations around the world. GBIF provides free and open access to biodiversity data from a wide range of sources, including scientific institutions, museums, and research projects.



1. Navigate to the occurrence search tool.
2. Search bar for text queries.

By default, when any text is entered into the search bar, the GBIF portal automatically initiates a search for all the data types. We suggest clicking on OCCURRENCES to restrict the search. Then it will initiate the **Occurrences** search, shown below.

SEARCH OCCURRENCES | 3,035,911,374 RESULTS

Scientific name	Country or area	Coordinates	Event date	Occurrence status	Basis
<i>Acacia implexa</i> Benth.	Australia	34.3S, 148.4E	2024 Jan 12	Present	Human
<i>Usnea intermedia</i> (A.Massal.) Jatta	Austria	47.2N, 15.5E	2024 Jan 27	Present	Human
<i>Tiliqua scincoides</i> (White, 1790)	Australia	33.8S, 151.2E	2024 Jan 23	Present	Human
<i>Mareca strepera</i> (Linnaeus, 1758)	France	48.9N, 2.8E	2024 Jan 07	Present	Human
<i>Pseudechirus peregrinus</i> (Boddaert, 1785)	Australia	33.7S, 151.1E	2024 Jan 23	Present	Human
<i>Ondatra zibethicus</i> (Linnaeus, 1766)	Netherlands (Kingdom of...)	51.5N, 6.1E	2024 Jan 18	Present	Human
<i>Sitta europaea</i> Linnaeus, 1758	Denmark	55.5N, 11.9E	2024 Jan 28	Present	Human
<i>Pseudechirus peregrinus</i> (Boddaert, 1785)	Australia	33.7S, 151.1E	2024 Jan 01	Present	Human
<i>Anas superciliosa</i> J.F.Gmelin, 1789	Australia	30.4S, 153.1E	2024 Jan 06	Present	Human
<i>Prunella modularis</i> (Linnaeus, 1758)	Germany	49.2N, 7.2E	2024 Jan 11	Present	Human
<i>Callidemum</i> Blanchard, 1853	Australia	35.3S, 149.1E	2024 Jan 04	Present	Human
<i>Sphecotheres vieilloti</i> Vigors & Horsfield, 1827	Australia	28.6S, 153.4E	2024 Jan 20	Present	Human
<i>Tiliqua scincoides</i> (White, 1790)	Australia	33.7S, 151.3E	2024 Jan 25	Present	Human

1. Text search bar
2. Filter Menu
3. Data retrieved visualization
4. Download button
5. Login button

In this example we will query for finger millet (*Eleusine coracana*) accession data, we strongly recommend using the scientific name as query, then click on the magnified glass icon.

The screenshot shows the GBIF Occurrences search results for the species "Eleusine coracana". A modal dialog is open asking if the user wants to limit the search to this taxon only. The "YES" button is highlighted with a red box.

Scientific name	Country or area	Coordinates	Event date	Occurrence status
<i>Eleusine coracana</i> (L.) Gaertn.	Sweden	57.8N, 12.0E	2004 Sep 25	Present
<i>Eleusine coracana</i> (L.) Gaertn.	Sweden	57.8N, 12.0E	2004 Sep 25	Present
<i>Eleusine coracana</i> (L.) Gaertn.	Sweden	56.7N, 12.9E	2012 Sep 28	Present
<i>Eleusine coracana</i> (L.) Gaertn.	Sweden	57.8N, 12.0E	2004 Sep 16	Present
<i>Eleusine coracana</i> (L.) Gaertn.	Sweden	57.8N, 12.0E	2004 Sep 25	Present
<i>Eleusine coracana</i> (L.) Gaertn.	Sweden	57.8N, 12.0E	2004 Sep 25	Present
<i>Eleusine coracana</i> (L.) Gaertn.	Sweden	57.8N, 12.0E	2004 Sep 11	Present
<i>Eleusine coracana</i> (L.) Gaertn.	Sweden	57.8N, 12.0E	2004 Sep 25	Present
<i>Eleusine coracana</i> (L.) Gaertn.	Netherlands (Kingdom of ...	51.1N, 5.9E	2024 Sep 20	Present
<i>Eleusine coracana</i> (L.) Gaertn.			1899 Dec 30	Present
<i>Eleusine coracana</i> (L.) Gaertn.	United Kingdom of Great ...	51.5N, 2.7W	1930 Sep 28	Present
<i>Eleusine coracana</i> (L.) Gaertn.	Germany	51.9N, 12.2E		Present
<i>Eleusine coracana</i> (L.) Gaertn.	Netherlands (Kingdom of ...	51.1N, 5.9E	2024 Sep 20	Present

Click the YES button to limit the results to accessions with an exact match. The retrieved data includes all synonyms of the crop genus. For example, if the filter criteria are “**Eleusine coracana**”, the results may also include “**Eleusine coracana subsp. coracana**”, which is considered a synonym of “**Eleusine coracana**”.

<i>Eleusine coracana</i> (L.) Gaertn.	Ukraine	50.4N, 30.6E	2014 Aug 13	Present	Human
<i>Eleusine coracana</i> (L.) Gaertn.	Kenya		2014 Aug 06	Present	Preserved
<i>Eleusine coracana</i> subsp. <i>coracana</i>	India		2014	Present	Materiel
<i>Eleusine coracana</i> subsp. <i>coracana</i>	India		2014	Present	Materiel

Once the final dataset is ready, you can download it by clicking the Download button. This will redirect you to the download page, where you'll need to log in with a valid GBIF account if you haven't already. Log-in options include Google and GitHub.

The screenshot shows the GBIF Occurrences search results for the species "Eleusine coracana". The "Simple filters" button is selected. The "Download" button is highlighted with a red box.

Format	Estimated data size
Tab-delimited CSV (for use in Excel, etc.) ⓘ	9 MB (2 MB zipped for download)
Tab-delimited CSV (for use in Excel, etc.) ⓘ	28 MB (6 MB zipped for download)
Tab-delimited CSV (for use in Excel, etc.) ⓘ	

Always download the “**DARWIN CORE ARCHIVE**” file. This will download a ZIP file with all the metadata related to the query. Within this zipped file you will find the next files:

dataset	File folder		
citations.txt	Text Document	3 KB	No
meta.xml	Microsoft Edge HTML Do...	4 KB	No
metadata.xml	Microsoft Edge HTML Do...	2 KB	No
multimedia.txt	Text Document	5 KB	No
occurrence.txt	Text Document	27 KB	No
rights.txt	Text Document	2 KB	No
verbatim.txt	Text Document	20 KB	No

The **occurrence.txt** file contains all the passport data for the downloaded accessions. Additional files provide supplementary information, such as **citations.txt**, which include details on how to cite the data for a scientific article.

## DOWNLOAD OCCURRENCES FROM USDA-GRIN

**NOTE:** This database contains occurrences for taxa available in the United States Department of Agriculture (This database only contains Germplasm bank accessions!).

In your browser type: <https://npgsweb.ars-grin.gov/gringlobal/search>. The following scheme will appear in your browser. The website possesses four tabs 1) Simple search (Search using PI number), 2) List search (This option serves to search a list of accessions (e.g. PI 651794,PI 651649,PI 651650) 3) Advanced search (This option is the only option that includes the search with species name 4) Results (This tab returns the search results). Also on the top of the website the scroll menu allows you to increase the number of accessions required by the user.

Select the tab for the type of search. Each tab has everything you need to do to perform that type of search.

Return up to 500 ▾

(Results of more than 500 will not return images.)

[Simple Search](#) [List Search](#) [Advanced Search](#) [Results](#)

---

e.g., PI 651650

This search will show only accessions that have material that may be requested today.

Please use the Advanced search and change the Return up to 10000 as follows to obtain an accurate number of accessions. Advanced search allows you to search by species name and limit the accession to countries if it is necessary. The following graphic shows how to search for records of *Amaranthus caudatus*. The search also will be limited to searching active accessions by activating the option Accessions that can be requested (including out of season). Finally, to search click on the search button.

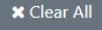
Select the tab for the type of search. Each tab has everything you need to do to perform that type of search.

**Return up to**  **Update Limit**

(Results of more than 500 will not return images.)

**Simple Search** **List Search** **Advanced Search** **Results**

**The more information you provide, the better the search will be.**

Scientific name (any part, no hybrid symbols)

Plant name

Genebank

Country of Origin   
Albania  
Algeria  
Angola  
...

**Reset Countries**

**Other search criteria:**

**Search for:**

- Accessions that can be requested (including out of season)  
 All accessions - Including historic (not in the NPGS collections, information only)

**Limit accessions displayed:**

- With genomic data  With NCBI link  With images  Only non-Genetically Engineered



The results will appear in the Results tab as shown in the following graphics. Please click on the View Accession Details to add the information attached to each germplasm passport and show/hide columns to show the new columns.

Select the tab for the type of search. Each tab has everything you need to do to perform that type of search.

Return up to  Update Limit

(Results of more than 500 will not return images.)

Simple Search List Search Advanced Search Results

If your results aren't what you expected, try using the Advanced Search tab and filling in more information.

Your query included: All accessions

Scientific name (any part, no hybrid symbols): Amaranthus caudatus

View Observation Data

Selected item(s) below:

Search: <input type="text"/>						
<input type="button" value="Basic Info"/>	<input type="button" value="Source Info"/>	<input type="button" value="Show all columns"/>	<input type="button" value="Show/hide columns"/>	<input type="button" value="Show 10 rows"/>	<input type="button" value="Excel"/>	
Showing 1 to 10 of 763 entries						
<input type="checkbox"/> ▲ ACCESSION	▼ PLANT NAME	▼ TAXONOMY	▼ ORIGIN	▼ GENE BANK	▼ AVAILABILITY	▼ IMPROVEMENT LEVEL
<input type="checkbox"/> PI 654390	DB 2005652	<i>Amaranthus caudatus</i> L.	Iowa, United States	NC7		Genetic material
<input type="checkbox"/> PI 632249	DB 2001949	<i>Amaranthus caudatus</i> L.	Iowa, United States	NC7		Genetic material
<input type="checkbox"/> PI 628794	Ames 23921	<i>Amaranthus caudatus</i> L.	Lima, Peru	NC7		Cultivar
<input type="checkbox"/> PI 669853	UC122	<i>Amaranthus caudatus</i> L.	Uttar Pradesh, India	NC7		Landrace
<input type="checkbox"/> PI 669855	UC124	<i>Amaranthus caudatus</i> L.	Uttar Pradesh, India	NC7		Landrace
<input type="checkbox"/> PI 669883	UC153	<i>Amaranthus caudatus</i> L.	Himachal Pradesh, India	NC7		Landrace
<input type="checkbox"/> PI 669885	UC158	<i>Amaranthus caudatus</i> L.	Uttar Pradesh, India	NC7		Landrace
<input type="checkbox"/> PI 669888	UC176	<i>Amaranthus caudatus</i> L.	Uttar Pradesh, India	NC7		Landrace
<input type="checkbox"/> PI 669889	UC177	<i>Amaranthus caudatus</i> L.	Uttar Pradesh, India	NC7		Landrace
<input type="checkbox"/> PI 669891	UC185	<i>Amaranthus caudatus</i> L.	Uttar Pradesh, India	NC7		Landrace
Showing 1 to 10 of 763 entries						
Previous	1	2	3	4	5	... 77 Next

The search had 763 entries. The columns possible to evaluate show accession number, taxonomy, origin, geneank, etc...Please download by clicking on the button Excel

<input type="button" value="Basic Info"/> <input type="button" value="Source Info"/> <input type="button" value="Show all columns"/> <input type="button" value="Show/hide columns"/> <input type="button" value="Show 10 rows"/> <input style="border: 2px solid red; padding: 2px; margin-left: 10px;" type="button" value="Excel"/> <input style="width: 150px;" type="text" value="Search: "/>										
<span style="margin-right: 10px;">Previous</span> <span style="border: 1px solid #ccc; padding: 2px 5px; background-color: #f0f0f0;">1</span> <span style="margin-left: 10px;">2</span> <span style="margin-left: 10px;">3</span> <span style="margin-left: 10px;">4</span> <span style="margin-left: 10px;">5</span> ... <span style="margin-left: 10px;">77</span> <span style="margin-left: 10px;">Next</span>										
<input type="checkbox"/> <span style="font-size: 1.5em;">▲</span> <span style="font-size: 0.8em;">ACCESSION</span>	<span style="font-size: 0.8em;">PLANT NAME</span>	<span style="font-size: 0.8em;">TAXONOMY</span>	<span style="font-size: 0.8em;">ORIGIN</span>	<span style="font-size: 0.8em;">GENEBANK</span>	<span style="font-size: 0.8em;">AVAILABILITY</span>	<span style="font-size: 0.8em;">RECEIVED</span>	<span style="font-size: 0.8em;">SOURCE TYPE</span>	<span style="font-size: 0.8em;">SOURCE DATE</span>	<span style="font-size: 0.8em;">COLLECTION SITE</span>	
<input type="button" value="Search ACC"/> <input type="button" value="Search"/> <input type="button" value="Search TAXC"/>										
<span style="color: green;">+</span>	<input type="checkbox"/> PI 654390	DB 2005652	<i>Amaranthus caudatus</i> L.	Iowa, United States	NC7		2007	DEVELOPED	05/06/2005	
<span style="color: green;">+</span>	<input type="checkbox"/> PI 632249	DB 2001949	<i>Amaranthus caudatus</i> L.	Iowa, United States	NC7		2002	DEVELOPED	01/24/2001	
<span style="color: green;">+</span>	<input type="checkbox"/> PI 628794	Ames 23921	<i>Amaranthus caudatus</i> L.	Lima, Peru	NC7		1997	DONATED	08/08/1997	
<span style="color: green;">+</span>	<input type="checkbox"/> PI 669853	UC122	<i>Amaranthus caudatus</i> L.	Uttar Pradesh, India	NC7		1996	COLLECTED	1978	Pauri to Badrinath transect.
<span style="color: green;">+</span>	<input type="checkbox"/> PI 669855	UC124	<i>Amaranthus caudatus</i> L.	Uttar Pradesh, India	NC7		1996	COLLECTED	1978	Pauri.
<span style="color: green;">+</span>	<input type="checkbox"/> PI 669883	UC153	<i>Amaranthus caudatus</i> L.	Himachal Pradesh, India	NC7		1996	COLLECTED	1978	Dashog.
<span style="color: green;">+</span>	<input type="checkbox"/> PI 669885	UC158	<i>Amaranthus caudatus</i> L.	Uttar Pradesh, India	NC7		1996	COLLECTED	1978	Joshimath.
<span style="color: green;">+</span>	<input type="checkbox"/> PI 669888	UC176	<i>Amaranthus caudatus</i> L.	Uttar Pradesh, India	NC7		1996	COLLECTED	1978	Pando Keshawar.
<span style="color: green;">+</span>	<input type="checkbox"/> PI 669889	UC177	<i>Amaranthus caudatus</i> L.	Uttar Pradesh, India	NC7		1996	COLLECTED	1978	Rampur.
<span style="color: green;">+</span>	<input type="checkbox"/> PI 669891	UC185	<i>Amaranthus caudatus</i> L.	Uttar Pradesh, India	NC7		1996	COLLECTED	1978	Rampur.

Showing 1 to 10 of 763 entries

Previous 1 2 3 4 5 ... 77 Next

The file downloaded is named Search Accessions GRIN-Global.xlsx and it looks like this. Please go the table below for technical details of the download.

ACCESSION	PLANT NAME	TAXONOMY	ORIGIN	GENDER	AVAILABILITY	RECOLLECTED	SOURCE TYPE	SOURCE DATE	COLLECTION SITE	COORDINATES	ELEVATE	HABITAT	IMPROVEMENT LEVEL	MATERIAL
P102430	CB 2005652	<i>Amaranthus caudatus</i> L.	Iowa, United States	MC7	2007	DEVELOPED	2005/06/05	2005/06/05	PI 654390	30 1400000.00 97 7150000.00	700	Landrace	Genetic material	Naturally occurring material. Grown as a accesso...
P102431	CB 2005653	<i>Amaranthus caudatus</i> L.	Iowa, United States	MC7	1997	COLLECTED	1997/01/01	1997/01/01	PI 632249	30 1400000.00 97 7150000.00	700	Landrace	Original material	Original material
P102432	CB 2005654	<i>Amaranthus caudatus</i> L.	Iowa, United States	MC7	1997	COLLECTED	1997/01/01	1997/01/01	PI 628794	30 1400000.00 97 7150000.00	700	Landrace	Original material	Original material
P102433	CB 2005655	<i>Amaranthus caudatus</i> L.	Iowa, United States	MC7	1997	COLLECTED	1997/01/01	1997/01/01	PI 669853	30 1400000.00 97 7150000.00	700	Landrace	Original material	Original material
P102434	CB 2005656	<i>Amaranthus caudatus</i> L.	Iowa, United States	MC7	1997	COLLECTED	1997/01/01	1997/01/01	PI 669855	30 1400000.00 97 7150000.00	700	Landrace	Original material	Original material
P102435	CB 2005657	<i>Amaranthus caudatus</i> L.	Iowa, United States	MC7	1997	COLLECTED	1997/01/01	1997/01/01	PI 669883	30 1400000.00 97 7150000.00	700	Landrace	Original material	Original material
P102436	CB 2005658	<i>Amaranthus caudatus</i> L.	Iowa, United States	MC7	1997	COLLECTED	1997/01/01	1997/01/01	PI 669885	30 1400000.00 97 7150000.00	700	Landrace	Original material	Original material
P102437	CB 2005659	<i>Amaranthus caudatus</i> L.	Iowa, United States	MC7	1997	COLLECTED	1997/01/01	1997/01/01	PI 669888	30 1400000.00 97 7150000.00	700	Landrace	Original material	Original material
P102438	CB 2005660	<i>Amaranthus caudatus</i> L.	Iowa, United States	MC7	1997	COLLECTED	1997/01/01	1997/01/01	PI 669889	30 1400000.00 97 7150000.00	700	Landrace	Original material	Original material
P102439	CB 2005661	<i>Amaranthus caudatus</i> L.	Iowa, United States	MC7	1997	COLLECTED	1997/01/01	1997/01/01	PI 669891	30 1400000.00 97 7150000.00	700	Landrace	Original material	Original material
P102440	CB 2005662	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	Not Available	1994	COLLECTED	09/04/1993	PI 669892	0.00000000 78 2050000.00	2050	Landrace	Wild and weedy	Wild and weedy plants and green dryings from the Lower...
P102441	CB 2005663	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669893	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102442	CB 2005664	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669894	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102443	CB 2005665	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669895	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102444	CB 2005666	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669896	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102445	CB 2005667	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669897	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102446	CB 2005668	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669898	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102447	CB 2005669	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669899	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102448	CB 2005670	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669900	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102449	CB 2005671	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669901	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102450	CB 2005672	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669902	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102451	CB 2005673	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669903	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102452	CB 2005674	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669904	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102453	CB 2005675	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669905	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102454	CB 2005676	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669906	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102455	CB 2005677	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669907	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102456	CB 2005678	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669908	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102457	CB 2005679	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669909	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102458	CB 2005680	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669910	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102459	CB 2005681	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669911	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102460	CB 2005682	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669912	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102461	CB 2005683	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669913	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102462	CB 2005684	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669914	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102463	CB 2005685	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669915	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102464	CB 2005686	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669916	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102465	CB 2005687	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669917	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102466	CB 2005688	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669918	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102467	CB 2005689	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669919	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102468	CB 2005690	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669920	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102469	CB 2005691	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669921	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102470	CB 2005692	<i>Amaranthus caudatus</i> L.	India, Puducherry	MC7	1990	COLLECTED	1990/01/01	1990/01/01	PI 669922	-21 4500000.00 -64 000000.00	2050	Landrace	Original material	The seeds are black.
P102471	CB 2005693	<i>A</i>												

**Table 1. Technical description of the fields available in USDA download.**

field	description	example
ACCESSION	Accession name	PI 634914
PLANT NAME	variety name	Ghanair
TAXONOMY	Species or taxon name	Amaranthus caudatus L.
ORIGIN	Locality information	Azad Jammu and Kashmir, Pakistan
GENEBANK	FAO WIEWS genebank id	NC7
AVAILABILITY	Availability in USDA	
RECEIVED	Received data	1989
SOURCE TYPE	Source	COLLECTED
SOURCE DATE	Date	10/25/1988
COLLECTION SITE	Collection Site data	Below road at Bandi Bakhalan, 5 km north of Chikar.
COORDINATES	decimal coordinates	34.15000000, 73.68333333
ELEVATION	Elevation data	1540
HABITAT	Habitat	Terraced farm. On a 5 degree slope facing northwest.
IMPROVEMENT LEVEL	type of record: CWR, landrace, or plant breeding	Landrace
NARRATIVE	Accession description	A grain type, with black seeds, red foliage and a drooping, red seed head. Used as a vegetable. As reported by the collectors. The seeds are black, flowers dark pink, leaves green. The RRC class type is: South American. In the greenhouse the leaves turned reddish as the plant matured. Observations from the Rodale Research Center, 1988 Rodale Amaranth Germplasm Catalog. Emmaus, PA.
	Accession unique id in USDA system	1051840

## **DOWNLOAD OCCURRENCES FROM FAO-WIEWS ([link](#))**

WIEWS is the information system used by FAO for the preparation of periodic, country-driven global assessments of the status of conservation and use of PGRFA. WIEWS also monitors, on the basis of country reports, the implementation of the Second Global Plan of Action for Plant Genetic Resources for Food and Agriculture, adopted in 2011.

## WIEWS - World Information and Early Warning System on Plant Genetic Resources for Food and Agriculture

[Home](#) [Background](#) [Data](#) [Resources](#) [Glossary](#)

### Ex situ search

Accession-level information of plant genetic resources secured in genebanks (*ex situ*) under medium and long term storage can be retrieved through the search below.

Year	Country
2022	Country

Holding institute

Holding institute	<input type="button"/> <input type="button"/>
Current selected institute(s)	<input type="button"/>
Selection currently empty	

Crop

1 Crop	2	Crop Wild Relatives
Genus	Species	<input type="button"/> <input type="button"/>
Current selected element(s)	3	<input type="button"/>
Selection currently empty		

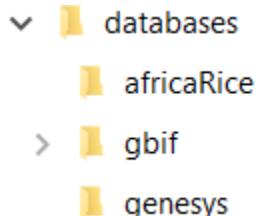
Among reported genera and species  Among accepted genera and species

Accession number	DOI
Country of origin	Status under the Multilateral System
5 Biological status of accession	Type of germplasm storage
Biological status of accession	<input type="button"/> <input type="button"/> 6

1. Genus's filter: type genus, automatically web portal will display the available genus (if nothing is displayed that means no data available for it).
2. Species filter: type one or more species, available species will be displayed (if nothing is displayed that means no data available).
3. Click to add genus and specie to the query
4. Displays the current selection of specie (user can remove those species that do not match the query)
5. Select option “300) Traditional cultivar/Landrace”
6. Click on search
7. Download resulting table

## PROTOCOL FOR DATABASES CONSOLIDATION - GAP ANALYSIS

1. Download each database and save them in separated folders, we encourage to name folders with the same name as the data source.



2. Apply the next filters for each one of the passport data we have downloaded (we encourage users to save this new filtered data table in a new file, preserving always the raw file):

- a. **GBIF:**

- i. Make sure that in the species column only appears the specie we wanted.
- ii. Remove rows with missing decimallatitude or decimallongitude.
- iii. Keep rows that matches "PRESERVED\_SPECIMEN" and "LIVING\_SPECIMEN" from **basisOfRecord** column. Remove any row that does not match the criteria.
- iv. Keep rows that match "ACCEPTED" and "SYNONYM" from **taxonomicStatus** column. Remove any row that does not match the criteria.
- v. Create a new column named "**status**" from the **basisOfRecord** following the next rules: set **G** if **basisOfRecord** == "LIVING\_SPECIMEN", set **H** in any other case.
- vi. Create a new column named "**ISO3**". From column **countryCode**, convert the existing ISO2 country code to ISO3 format.
- vii. Create a new column named "**COLLSITE**". Concatenate columns **locality**, **municipality**, **country** and **stateProvince** into one single character string.
- viii. Create a new column named "**source\_db**". This column should have the name of the data source where it was downloaded.
- ix. Create a new column named "**database\_id**". This column should have a consecutive sequence of numbers that identifies each row from the data table.
- x. Keep the next columns: **ISO3**, **elevation**, **institutionCode**, **decimalLatitude**, **decimalLongitude**, **catalogNumber**, **COLLSITE**, **status**, **source\_db**, **database\_id**.
- xi. Rename columns as follows:  
**ORIGCTY** = ISO3,  
**ELEVATION** = elevation,  
**INSTCODE** = institutionCode,  
**DECLATITUDE** = decimalLatitude,  
**DECLONGITUDE** = decimalLongitude,  
**ACCENUMB** = catalogNumber

- b. **Gensys:**

- i. Create a new column named **status**, fill this column with a fixed lab "**G**".
- ii. Create a new column named **source\_db**, fill this column with a fixed lab "**GENESYS**".
- iii. Create a new column named **database\_id**, This column should have a consecutive sequence of numbers that identifies each row from the data table.

- iv. Remove rows with missing **DECLATITUDE** or **DECLONGITUDE**.
  - v. Select rows that match “**300**” from **SAMPSTAT**, remove rows that does not match.
  - vi. Keep the next columns: "**SAMPSTAT**", "**ORIGCTY**", "**ELEVATION**", "**DECLATITUDE**", "**DECLONGITUDE**", "**ACCENUMB**" and "**INSTCODE**".
3. Once you have the databases, please check that they only have records of the target race, gender or class. In case of GBIF, for example, the database usually has records of the same crop but these are not always landraces or have the same class that you are looking for:

gbifid	species	infraspeci	taxonrank	countryco	locality	decimalla	decimallo	elevation	basisofrec	institution	address	lon.geoco	lat.geoco
910479696	Oryza sativa	NA	SPECIES	ES	Guadiana	38.95932	-6.63416	NA	PRESERVE	HSS	NA	NA	NA
910479701	Oryza sativa	NA	SPECIES	ES	Guadiana	38.95932	-6.63416	NA	PRESERVE	HSS	NA	NA	NA
912100503	Oryza sativa	NA	SPECIES	IN	Raigarh	NA	NA	NA	PRESERVE	K	Raigarh ,IN	73.43333	18.25
912295773	Oryza glaberrima	NA	SPECIES	ML	French Su	NA	NA	NA	PRESERVE	K	French Su	-3.25756	14.77709
912295802	Oryza glaberrima	NA	SPECIES	SD	Casamanc	NA	NA	NA	PRESERVE	K	Casamanc	-16.7592	12.55611

In this case, although all the Oryza records were downloaded, we want to analyze just **glaberrima**, that is why every record different to *Oryza glaberrima* must be removed. The filtered database must be saved as a new file called “**<db\_name>\_cleaned.csv**” and the original database **MUST BE KEPT WITHOUT MODIFICATIONS OR FILTERS**.

4. **[IMPORTANT]** Add two new columns in each database:
- **source\_db**: Data source name, i.e. Genesys, gbif, CIAT, etc.
  - **database\_id**: Identifier of the records in each database. It has an abbreviator of the data source name, then an underscore “**\_**”, and a unique number for each record. The unique number can be the “accession number”.

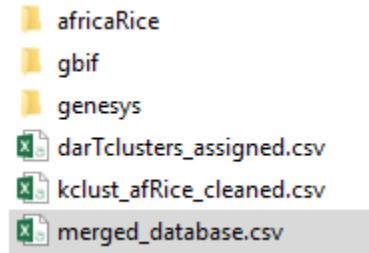
ecology	dartsec	dartsec	declat	declon	MAIN	Total.n	Number	Plant.h	Panicle	Leaf.bl	Leaf.bl	status	source_db	database_id
NA	NA	NA	14.77709	-3.25756	NA	NA	NA	NA	NA	NA	NA	H	gbif	gbif_1
NA	NA	NA	12.55611	-16.7592	NA	NA	NA	NA	NA	NA	NA	H	gbif	gbif_2
NA	NA	NA	10.9552	-2.6	NA	NA	NA	NA	NA	NA	NA	G	gbif	gbif_3
NA	NA	NA	12.5	4.333333	NA	NA	NA	NA	NA	NA	NA	G	gbif	gbif_4
NA	NA	NA	12.32528	-16.5478	NA	NA	NA	NA	NA	NA	NA	G	gbif	gbif_5
NA	NA	NA	11.38333	5.5	NA	NA	NA	NA	NA	NA	NA	G	gbif	gbif_6
NA	NA	NA	8.133333	-10.8833	NA	NA	NA	NA	NA	NA	NA	G	gbif	gbif_7
NA	NA	NA	11.16667	4.666667	NA	NA	NA	NA	NA	NA	NA	G	gbif	gbif_8
NA	NA	NA	10.4	14.85	NA	NA	NA	NA	NA	NA	NA	G	gbif	gbif_9
NA	NA	NA	8.766667	-10.5	NA	NA	NA	NA	NA	NA	NA	G	gbif	gbif_10

5. Do not remove the columns for each downloaded database. As you know, GBIF database has a huge amount of variables that are usually unnecessary, however, do not remove them.

gbifID	datasetKey	occurrenceID	Kingdom	phylum	class	order	family	genus	species
1840510370	b740eaa0-0679	http://www.botanicus.org/taxon/1840510370	Plantae	Tracheophyta	Liliopsida	Poales	Poaceae	Oryza	Oryza sativa
1840516273	b740eaa0-0679	http://www.botanicus.org/taxon/1840516273	Plantae	Tracheophyta	Liliopsida	Poales	Poaceae	Oryza	Oryza sativa
1840516392	b740eaa0-0679	http://www.botanicus.org/taxon/1840516392	Plantae	Tracheophyta	Liliopsida	Poales	Poaceae	Oryza	Oryza sativa

belong to GBIF. However, in order to determine if the accession is **G** or **H**, please follow the next:

- a. Filter by **basisOfRecord** the accession different from "MATERIAL\_SAMPLE" and "FOSSIL\_SPECIMEN".
  - b. Add a new field/variable/column called **status** to the database by Filtering **basisOfRecord**; if this variable is equal to "LIVING\_SPECIMEN" for a specific record, then the **status** of that accession must be **G**, otherwise **H**.
  - c. Make sure to preserve the column catalogNumber.
7. Merge the databases in one file, which must be saved in:  
`<root_path>/input_data/by_crop/rice_african/databases`. Please include the word "**merged**" in the file name, in order to identify the final database.



In the above example, the file **merged\_database.csv** is the final database created by the union of three different data sources: AfricaRice, GBIF and Genesys. In addition, there are two databases created from the final one: darTclusters\_assigned and kclust\_afRice\_cleaned in which different African rice genetic structures are proved, hence, it is easy to identify which is the database to do the analysis from which other databases are created.