

Stephan Weise, Matija Obreza



Use of the genebank metrics in aggregator information systems like EURISCO and Genesys

ECPGR Workshop on Genebank Metrics, 24-25 February 2026, Prague, Czech Republic

Overview

- Composition
 - Ten mandatory and 38 optional metrics
 - Five thematic groups
 - Size and composition of the collection
 - Data and documentation
 - Conservation
 - Availability
 - Distribution
- Can the metrics be calculated by aggregator systems?
 - Which ones?
 - To what extent?



Plant Genetic Resources Characterization and Utilization

cambridge.org/pgr

Research Article

Cite this article: van Hintum T et al. (2025) A proposal for genebank metrics to enhance collection management. *Plant Genetic Resources Characterization and Utilization* 1-11. <https://doi.org/10.1017/S147926212510021X>

Received: 4 April 2025
Revised: 18 July 2025
Accepted: 31 July 2025

Keywords:
ex situ conservation; genebank management; crop diversity; plant genetic resources; quality management system

Corresponding author: Theo van Hintum; Email: theo.vanhintum@wur.nl

A proposal for genebank metrics to enhance collection management

Theo van Hintum¹, Bela Bartha², Claudio Niggli², Alvina Avagyan³, Sylvia Vogl⁴, Lisa Achathaler⁴, Vojtech Holubec⁵, Ludmilla Papouškova⁶, Francesco Ferrari⁶, Graziano Rossi⁶, Attila Simon⁷, Lajos Horvath⁷, Renata Kowalik⁸, Maja Boczkowka⁸, Stephan Weise⁹, Markus Oppermann⁹, Maarten van Zonneveld¹⁰, Matija Obreza¹¹, Erik Wijnker¹, Noam Chayut¹², Fadia Chairi¹³, Johan Axelsson¹⁴, Lise Lykke Steffensen¹⁴ and Filippo Guzzon¹⁵

¹Centre for Genetic Resources, The Netherlands (CGN), Wageningen University and Research (WUR), Wageningen, The Netherlands; ²Department of Director, ProSpecieRara, Basel, Switzerland; ³Scientific Unit of Vegetable and Industrial Crops, Darakert, Armenia; ⁴Department for Plant Genetic Resources, Austrian Agency for Health and Food Safety, Linz, Austria; ⁵Genebank, Czech Agrifood Research Center, Prague, Czech Republic; ⁶Department of Earth and Environmental Sciences, University of Pavia, Pavia, Italy; ⁷National Centre for Biodiversity and Gene Conservation, Tüpfölszele, Hungary; ⁸Plant Breeding and Acclimatization Institute - National Research Institute (IHAR-PIB), National Centre for Plant Genetic Resources: Polish Genebank, Radzików, Poland; ⁹Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Germany; ¹⁰World Vegetable Center, Shanhua, Taiwan; ¹¹Information Systems Program, Global Crop Diversity Trust, Bonn, Germany; ¹²John Innes Centre, Norwich, UK; ¹³Wallon Centre of Agricultural Research, Libramont, Belgium; ¹⁴NordGen Plants, Nordic Genetic Resource Center (NordGen), Alnarp, Sweden and ¹⁵European Cooperative Programme for Plant Genetic Resources (ECPGR), c/o Alliance of Biodiversity International and CIAT, Rome, Italy

Abstract

Genebanks worldwide play a crucial role in the long-term conservation and accessibility of wide diversity of plant genetic resources (PGR). These resources are essential for addressing current and future challenges related to climate change and biodiversity loss, thereby contributing to agricultural production and sustainability. Enhancing the efficiency of genebanks requires implementing a quality management system that ensures data transparency, as well as the documentation and standardization of processes. This study proposes a set of 10 mandatory and 38 optional genebank metrics specifically designed for collections of orthodox seeds. These metrics define a set of basic parameters that are easy to calculate and collectively serve as a valuable tool for improving genebank management, enhancing performance, fostering transparency and promoting collaborations among institutions. Fourteen institutions engaged in the long-term conservation of PGR were consulted in the development of the proposed metrics. This work lays the foundation for the establishment of a widely accepted set of genebank metrics within the global genebank community. Further research and iterative development are necessary to formulate comprehensive metrics applicable to other *ex situ* conservation methods (such as cryopreservation, *in vitro* and field conservation) as well as *in situ* conservation strategies (including genetic reserves and on-farm conservation).

Introduction

The importance of conserving plant genetic resources (PGR) for future generations and their accessibility to contemporary plant scientists and breeders is widely acknowledged (King *et al.*, 2024). Genebanks, managing *ex situ* collections of PGR, play an essential role in enhancing agricultural production and sustainability (McCouch *et al.*, 2013). According to FAO (2025a), in 2022 the global genebank network conserved approximately 5.9 million accessions across 871 genebanks. Thirteen of these genebanks are international (such as those managed by the CGIAR), six are regional and the remaining 852 are national genebanks – more than half of them located in Europe. Based on the data associated with the material conserved in these genebanks, gap analyses can be conducted to identify areas where plant genetic diversity has been under-collected, enabling the establishment of targeted collecting missions and conservation programmes (Dulloo and Khoury, 2023). However, it is imperative to examine whether the material housed in genebanks is being adequately conserved and is readily accessible for utilization.

The performance of genebanks is a sensitive issue, with limited public data available regarding genebanks' efficacy in fulfilling their mandates. However, informal exchanges within

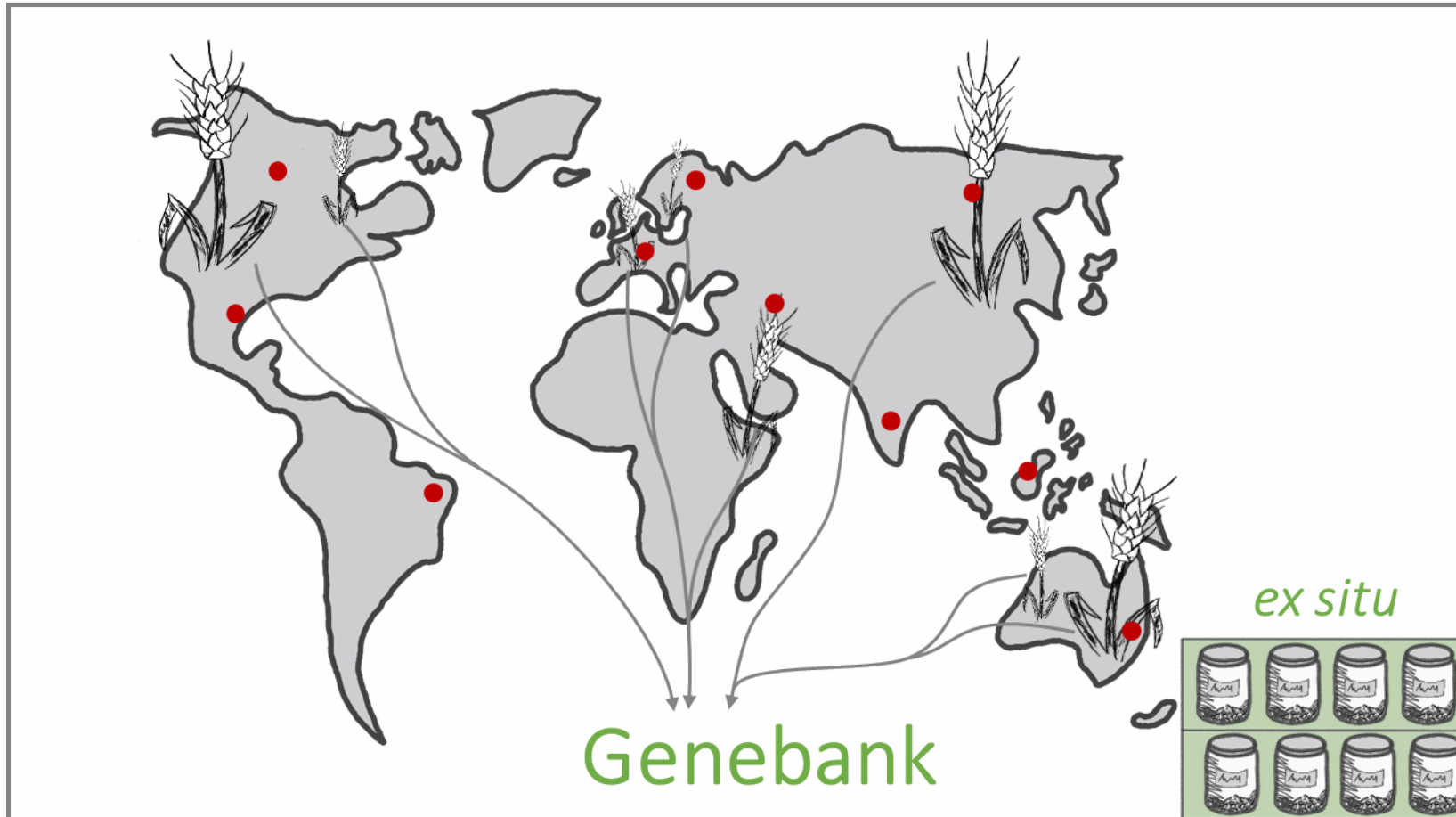
NIAB

© The Author(s), 2025. Published by Cambridge University Press on behalf of National Institute of Agricultural Botany. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Aggregator systems

- What are aggregator systems?
 - Integrate data from a large number of genebanks
 - Data gets refreshed periodically
- Why do we need them?
 - Provide a central entry point on information about PGR holdings
 - Even for collections that do not have their own information systems
 - Assist in meeting national obligations
- How many PGR accessions are we talking about?

Plant genetic resources collections



- 5.9 million *ex situ* accessions globally
- 13 international genebanks
- 4 regional genebanks
- 850 national genebanks

SoW3, 2025

- Hundreds of further PGR collections

EURISCO + Genesys



- EURISCO
 - <https://ecpgr.org/eurisco/>
 - Provides passport data and phenotypic data about plant germplasm accessions maintained in Europe and some neighbouring countries
 - European hub of Genesys
- Genesys
 - <https://www.genesys-pgr.org/>
 - International information system for plant genetic resources
 - Comprises EURISCO, CGIAR and additional large genebanks (USDA-GRIN, EMBRAPA, ...)



→ Both systems together document more than 4.4 million accessions

Metrics' application



- Basis of possible calculations: MCPD passport data
- Application to thematic metrics groups
 - Impossible to calculate with MCPD data
 - Conservation (except safety duplicates)
 - Availability
 - Distribution
 - Almost all metrics can be calculated
 - Size and composition of the collection
 - Data and documentation
 - Exceptions
 - DOC_GEN (Number of accessions with accessible genomic information)
 - NACC_OUT (Number of accessions removed from the collection)

→ Calculation only possible for data that has been made available

→ Timeliness and scope of data determined by providers

→ Demonstration using the example of EURISCO

The EURISCO logo, with the tagline 'Finding seeds for the future', is located at the top left of the document.

Descriptors for uploading passport data to EURISCO
2017-11-13

1 Introduction
This descriptor list describes the data exchange format for uploading passport data from the National Inventories to EURISCO.
The list is an extension of the FAO/Bioversity multi-crop passport descriptors (MCPD), version 2.1, which were published December 2015, developed jointly by Bioversity and FAO, with input from many documentation specialists worldwide, to provide international standards to facilitate germplasm passport information exchange. All MCPDs are included, without change and with the same formatting rules, in the current list. Four additional descriptors (00, 29-31) were added for the specific purposes of EURISCO.

2 General formatting rules
The following formatting rules, as copied from the MCPD documentation, apply to all fields:

- If a field allows multiple values, these values should be separated by a semicolon (;) without space (e.g. Accession name: Symphony;Emma;Songino).
- A field for which no value is available should be left empty (e.g. Elevation). If data are exchanged in ASCII format, a field with a missing numeric value should be left empty. If data are exchanged in a database format, missing numeric values should be represented by generic NULL values.
- Dates are recorded as YYYYMMDD. If the month or day are missing, this should be indicated with hyphens or '00' [double zero]. If both (month and day) are missing, two double zeros are needed (e.g. 1975---, 19750000; 197506--, 19750600).
- Country names: Three letter ISO codes are used for countries. The ISO 3166-1 standard country or area codes are available online at: <http://unstats.un.org/unsd/methodology/m49/>.
Note: The list of obsolete codes can be found at: http://en.wikipedia.org/wiki/ISO_3166-1_alpha-3#Reserved_code_elements.
- For institutes, the codes from FAO WIEWS should be used. The current set of institute codes is available from the FAO WIEWS site (<http://www.fao.org/wiews>).
 - If new institute codes are required, they can be generated online by FAO National Focal Points (<http://www.fao.org/agriculture/crops/thematic-sitemap/theme/seeds-pgr/national-focal-points/en/>) or they can be requested from: WIEWS@fao.org.
 - For institutes that no longer exist, or that were not assigned an FAO WIEWS institute code, please provide full details in descriptors 4.1, 4.1.1, 18.1, 22.1 and 25.1.

3 Descriptors
The descriptors are numbered according to the FAO/Bioversity multi-crop passport descriptors (MCPD), version 2.1. The first descriptor (numbered 00) and the last three (numbered 29-31) are additional, and specific to this EURISCO Descriptor List. Only the four fields identifying the accession are mandatory, all other fields are highly recommended. The mandatory fields are NICODE (00), INSTCODE (1), ACCENUMB (2) and GENUS (5). The combination of these fields has to be unique.

Size and composition of the collection

Size and composition of the collection

- Mandatory metric (as of 13 February 2026)
 - NACC (Number of accessions)
 - Total: 2,121,228
 - *Ex situ*: 2,114,186
 - *In situ* CWR: 7,042

- In the following, focus only on *ex situ* data

Country	Ex situ accs.	Country	Ex situ accs.
Albania	5222	Italy	56402
Armenia	7203	Latvia	2798
Austria	12071	Lithuania	2296
Azerbaijan	13430	Moldova	6015
Belarus	203	Montenegro	433
Belgium	11658	Netherlands	24217
Bosnia and Herzegovina	971	Nordic Countries	41307
Bulgaria	71401	North Macedonia	2158
Croatia	4879	Poland	90823
Cyprus	631	Portugal	30484
Czech Republic	59147	Romania	46378
Estonia	3732	Russian Federation	200717
France	23229	Serbia	6197
Georgia	440	Slovakia	18969
Germany	186392	Slovenia	3008
Greece	6284	Spain	80436
Hungary	50460	Switzerland	40016
Ireland	1711	Türkiye	12996
Israel	28884	Ukraine	112914
Italy	56402	United Kingdom	847674

Size and composition of the collection

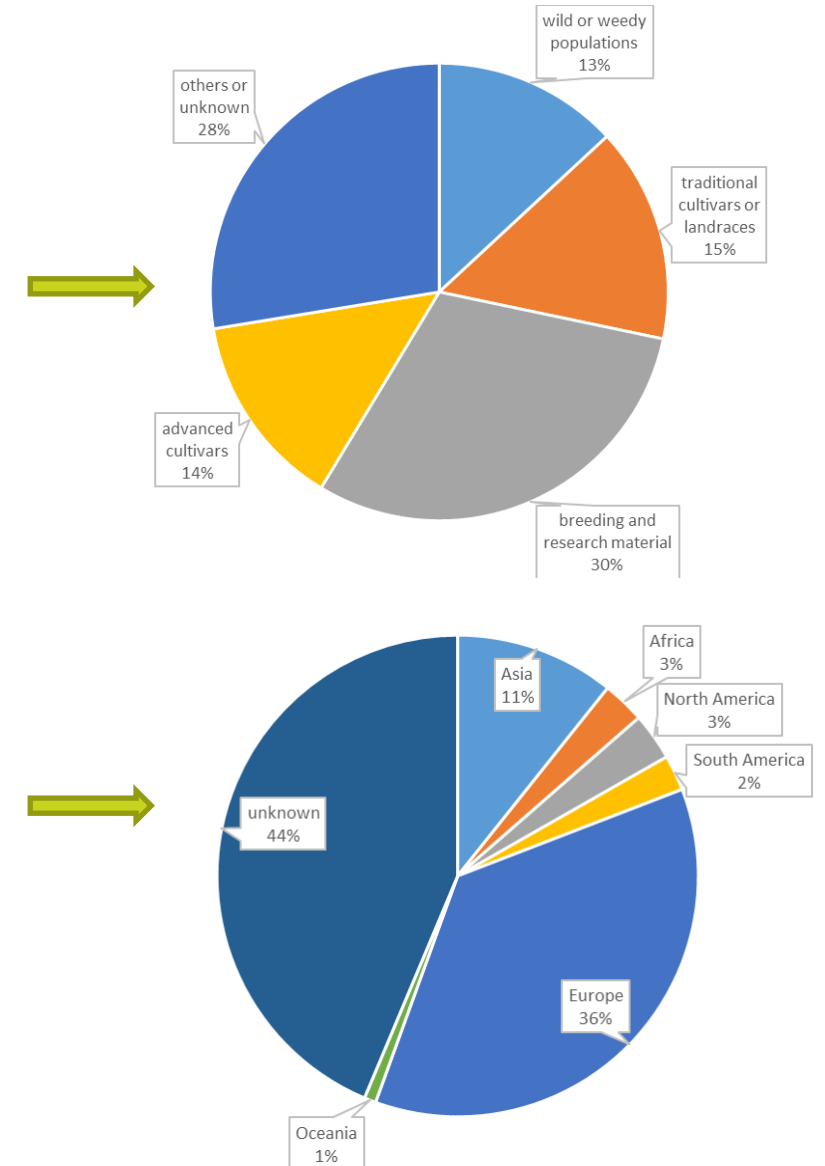
- Optional metrics

- Distinction by biological status

- NACC_PW (Number of accessions of wild or weedy populations): 276,855
- NACC_PL (Number of accessions of traditional cultivars or landraces): 321,332
- NACC_PB (Number of accessions of breeding and research material): 641,509
- NACC_PA (Number of accessions of advanced cultivars): 291,470

- Distinction by origin

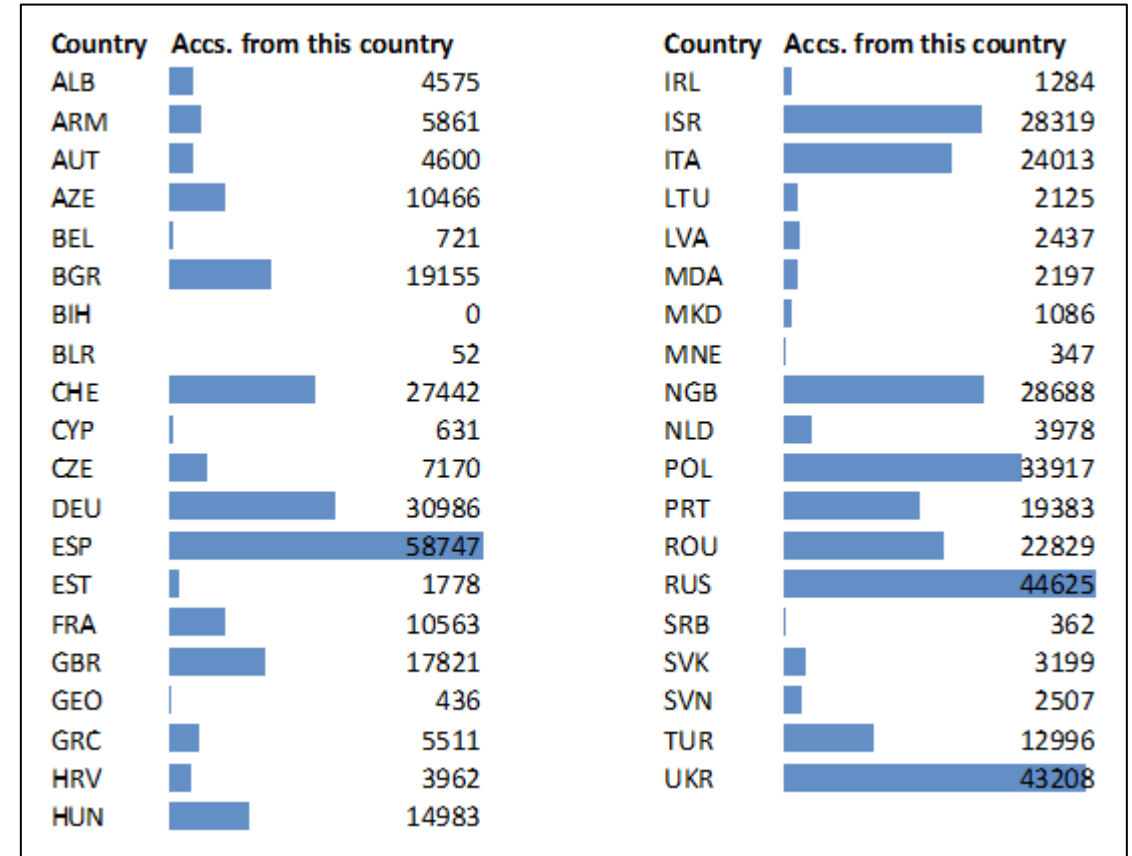
- NACC_OAS (Number of accessions originating in Asia): 226,819
- NACC_OAF (Number of accessions originating in Africa): 60,236
- NACC_ONA (Number of accessions originating in North America): 67,340
- NACC_OSA (Number of accessions originating in South America): 50,377
- NACC_OEU (Number of accessions originating in Europe): 769,513
- NACC_OAU (Number of accessions originating in Oceania): 16,683



Size and composition of the collection



- Optional metrics
 - Distinction by origin (cont.)
 - NACC_ON (Number of accessions originating in the country where the genebank is located): 502,960
 - Various challenges
 - DEU includes former East and West Germany → easy to query
 - NGB comprises DNK, SWE, NOR, FIN, ISL → easy to query
 - No ORIGCTY for BIH
 - Former Czechoslovakia divided
 - Former Yugoslavia divided



Size and composition of the collection

- Optional metrics

- NACC_NEW (Number of accessions added):

- 23,387
- ACQDATE unknown for 439,643 accessions
- Reporting period 1 January 2020 – 13 February 2026



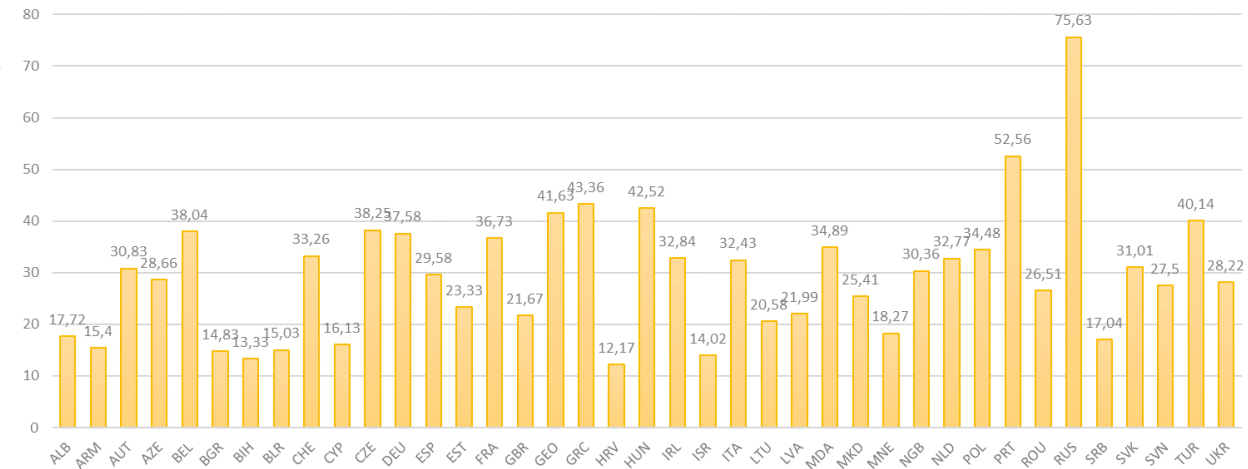
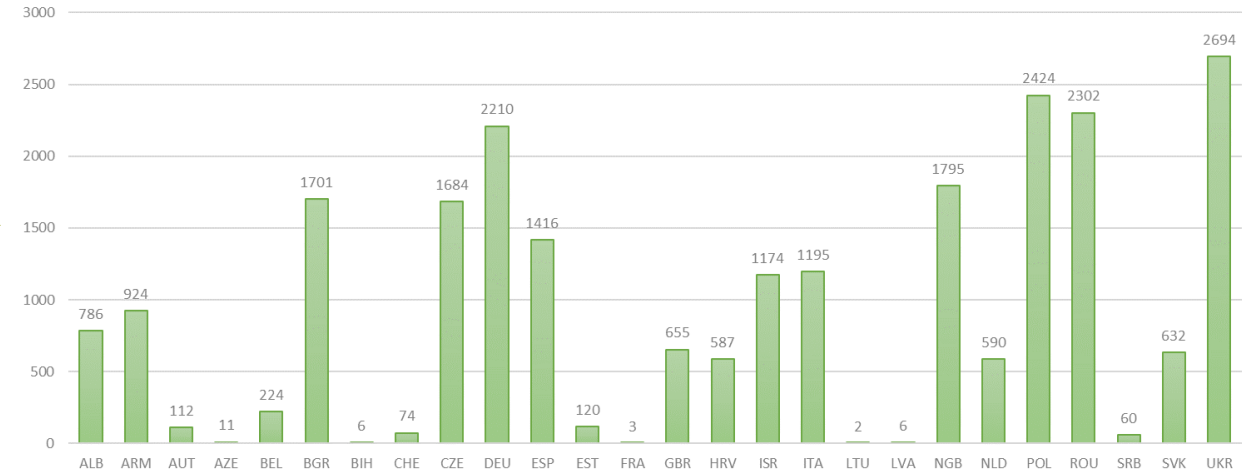
- ACC_AGE (Average time in years that the accessions have been in the collection):

- 27.98 years
- ACQDATE unknown for 439,643 accessions
- As of 13 February 2026



- NACC_OUT (Number of accessions removed from the collection):

- This information is not recored in EURISCO



Data and documentation

Data and documentation

- Mandatory metrics
 - DOC_PDCI (Average passport data completeness index (PDCI) of the accessions):
 - 5.33 (total)
 - What does it mean?

Clarification: Passport data completeness

- 45 passport data descriptors in EURISCO-MCPD
 - 4 mandatory
 - 41 optional
- Often, only limited information about certain accessions
- Some descriptors only sparsely populated
- In many cases, information available in the respective information systems, but not in EURISCO

National inventory
 National Inventory Code **BGR**
 National Inventory **Bulgaria**

Holding institute
How to get access to germplasm:
EURISCO does not provide the possibility to order accessions directly. The requests should be addressed to the holding institutions. For addresses and contact details you can check the [FAO-WIEWS](#) database. The ECPGR Secretariat (lmaggioli@cgiar.org) can also be contacted for further information.
 Institute Code **BGR001** ([Contact details on FAO-WIEWS website](#))
 Institute Name **Institute for Plant Genetic Resources 'K.Malkov', Sadovo, Plovdiv district, Bulgaria**

Accession
 Accession Number **1982-HOR-VU-84**
 Accession Names **Cernomorec**
 Crop Names **barley**
 MLS Status **unknown**
 AEGIS Status **unknown**

Taxonomy
 Genus **Hordeum**
 Species **vulgare**
 Species Authority **L.**
 Subtaxa **convar. distichon var. nutans**
 Subtaxa Authority **(Rode.) Alef.**

Acquisition/storage
 no data found

Collection
 no data found

Donor
 no data found

Breeder
 no data found

Other
 Other Numbers **BGR001:82105243**

Clarification: Completeness per descriptor (whole EURISCO)

Descriptor	Completeness [%]
GENUS	100
ACCENUMB	100
INSTCODE	100
NICODE	100
SPECIES	98.50
STORAGE	95.67
SPAUTHOR	92.22
CROPNAME	82.23
ACQDATE	79.19
SAMPSTAT	73.74
MLSSTAT	69.52
ACCEURL	57.66
COLLSRC	57.27
ORIGCTY	56.47
AEGISSTAT	54.30

Descriptor	Completeness [%]
DONORNUMB	45.97
ACCENAME	45.73
DUPLSITE	43.12
DONORNAME	42.17
BREDCODE	36.71
BREDNAME	34.97
DUPLINSTNAME	33.20
ANCEST	30.10
DONORCODE	25.79
COLLNUMB	21.69
COLLDATE	21.04
SUBTAXA	20.93
COLLSITE	19.44
COLLCODE	19.38
COLLNAME	14.49

Descriptor	Completeness [%]
OTHERNUMB	14.42
PUID	14.27
HISTORIC	13.40
ELEVATION	12.55
SUBTAUTHOR	12.27
LATITUDE	11.03
DECLATITUDE	10.97
LONGITUDE	10.94
DECLONGITUDE	10.91
REMARKS	6.62
COLLMISSID	4.43
COORDDATUM	1.58
COORDUNCERT	1.15
GEOREFMETH	1.00
COLLINSTADDRESS	0.49

Clarification: Passport Data Completeness Index (PDCI)

- Following van Hintum et al., Quality indicators for passport data in *ex situ* genebanks, *Plant Genetic Resources*, 9(3):478-485, 2011
→ adapted to EURISCO-MCPD v2
- Index between 0 (lowest quality) and 10 (highest quality)
- 60% of score independent of population type
- 40% of score depending on population type

Clarification: Passport Data Completeness Index (PDCI)

- Independent of population type

Descriptor	Score Condition	Remark
NICODE	0	mandatory
INSTCODE	0	mandatory
ACCENUMB	0	mandatory
GENUS	0	mandatory
PUID	40	
SPECIES	90 if GENUS is not null	
SPAUTHOR	5 if GENUS+SPECIES are not null	
SUBTAXA	45 if GENUS+SPECIES are not null	
SUBTAUTHOR	5 if GENUS+SPECIES+SUBTAXA are not null	
CROPNAME	45	
ACQDATE	30	
SAMPSTAT	80	
DONORCODE	50	
DONORNAME	0 or 20 if DONORCODE is null	
DONORNUMB	50 or 20 if DONORCODE is null	
OTHERNUMB	35	
DUPLSITE	30	
DUPLINSTNAME	0 or 15 if DUPLSITE is null	
STORAGE	20	
REMARKS	0	
ACCEURL	40	
MLSSTAT	15	
AEGISSTAT	15	
HISTORIC	5	

Clarification: Passport Data Completeness Index (PDCI)



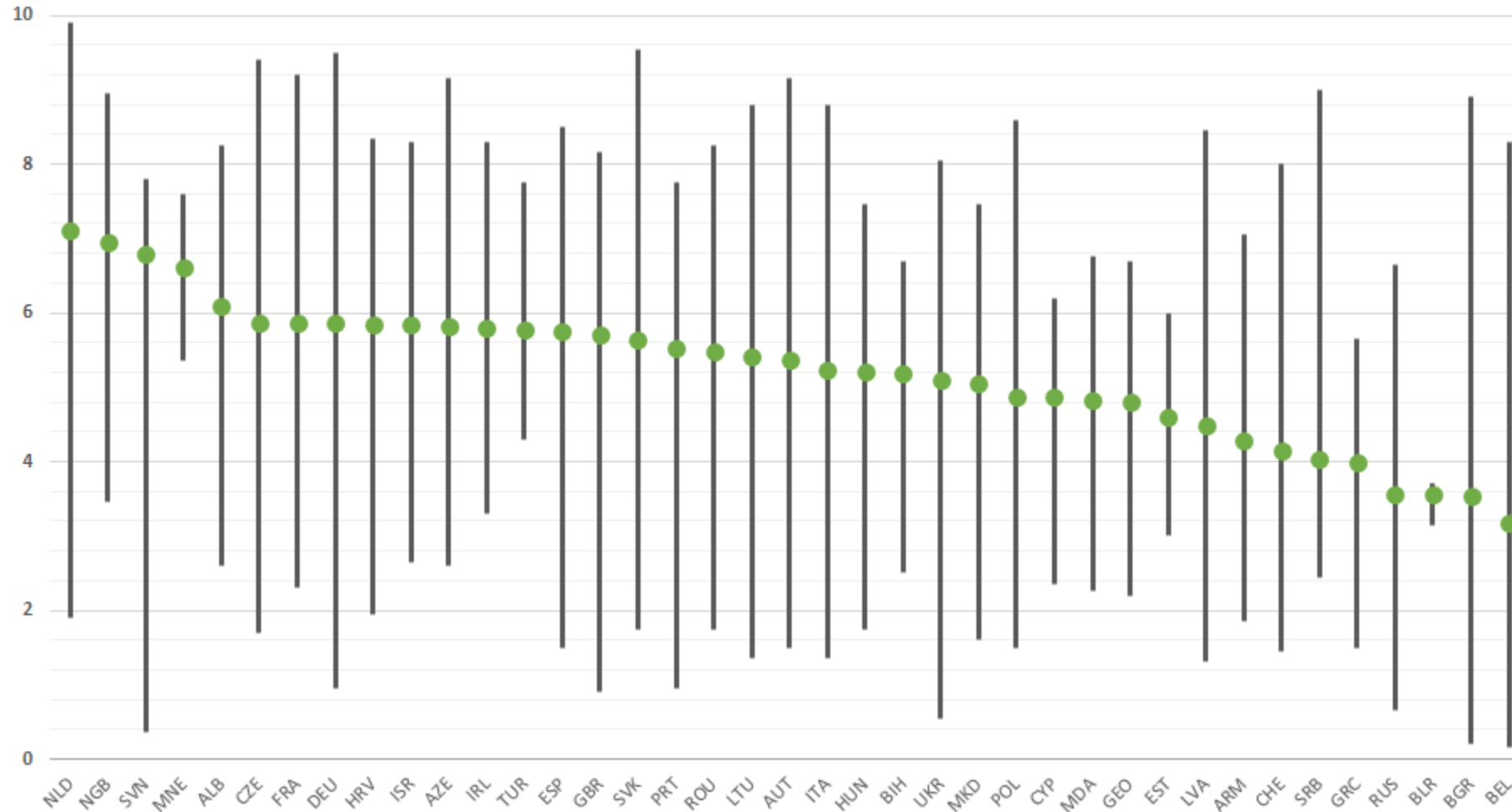
- Depending on population type

Descriptor	Wild/weedy	Landrace	Breeding/research material		Cultivar	Other/unknown
	Score Condition	Score Condition	Score	Condition	Score Condition	Score Condition
COLLNUMB	60	40		0	0	20
COLLCODE	40	30		0	0	20
COLLNAME	0 or 20 if COLLCODE null	0 or 15 if COLLCODE null		0	0	0 or 10 if COLLCODE null
ACCENAME	0	50	80		160	40
ORIGCTY	75	75	35		35	35
COLLSITE	20 or 70 if LATITUDE/LONGITUDE null	15 or 45 if LATITUDE/LONGITUDE null		0	0	10 or 20 if LATITUDE/LONGITUDE null
LATITUDE	60 if LONGITUDE not null	40 if LONGITUDE not null		0	0	15 if LONGITUDE not null
LONGITUDE	60 if LATITUDE not null	40 if LATITUDE not null		0	0	15 if LATITUDE not null
GEOREFMETH	5	5		0	0	5
ELEVATION	20	15		0	0	5
COLLDATE	30	30		0	0	10
BREDCODE	0	0	115		85	10
BREDNAME	0	0		0 or 55 if BREDCODE null	0 or 40 if BREDCODE null	0 or 10 if BREDCODE null
ANCEST	0	10	150		100	40
COLLSRC	30	50	20		20	25

The individual scores are added together and divided by 100.

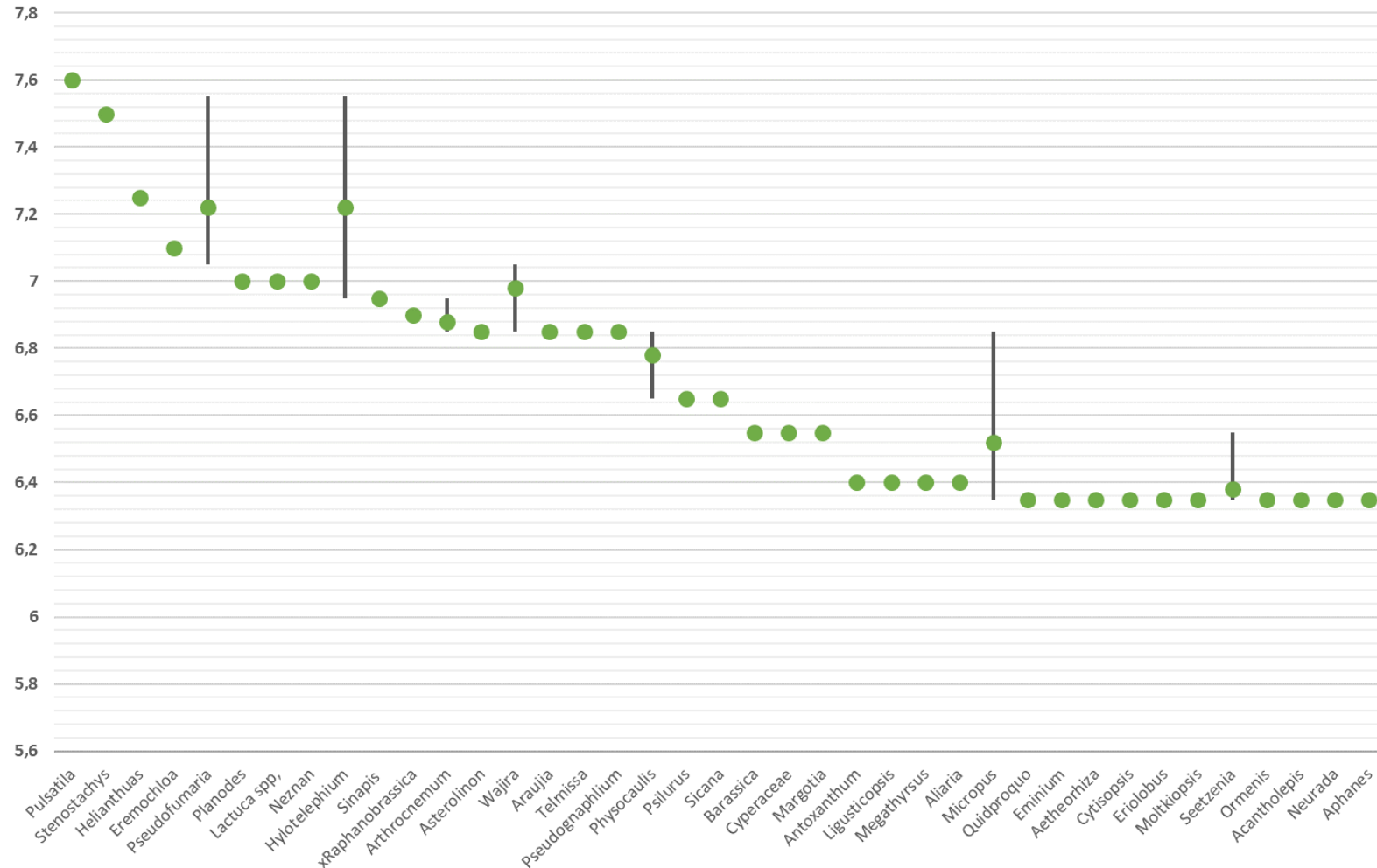
Passport Data Completeness Index (PDCI)

- Results per National Inventory



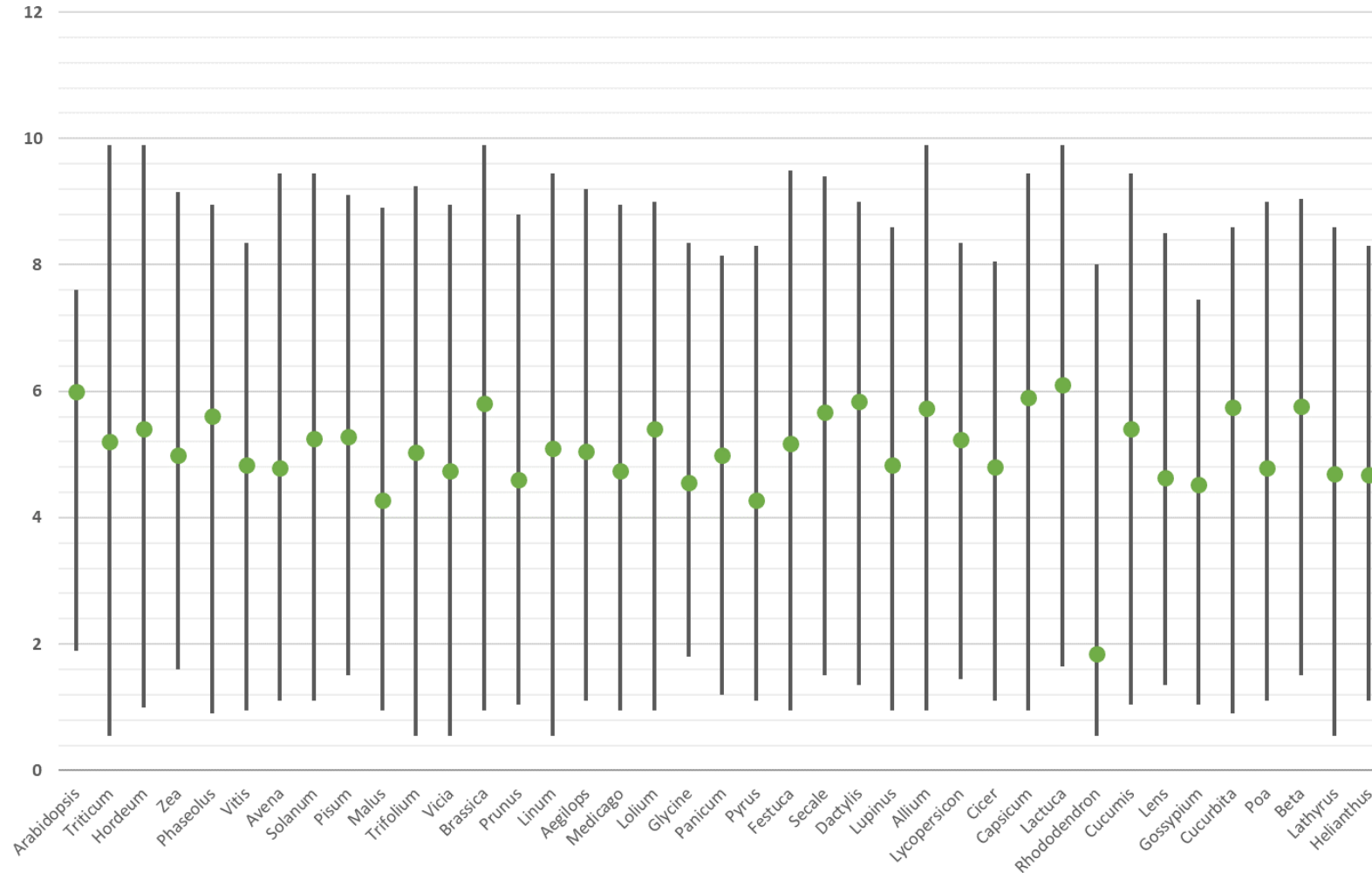
Passport Data Completeness Index (PDCI)

- Results per genus (for highest PDCI values only)



Passport Data Completeness Index (PDCI)

- Results per genus (for largest no. of accessions per genus)



Data and documentation

- **Mandatory metrics**

- **DOC_PDCI (Average passport data completeness index (PDCI) of the accessions):**

- 5.33 (total)
- Min-max: 0.15 ↔ 9.90 (total)

- **DOC_PHEN (Average number of phenotypic datapoints per accession):**

- 29.52
- Observation data for 92,907 accessions from 21 countries in database only

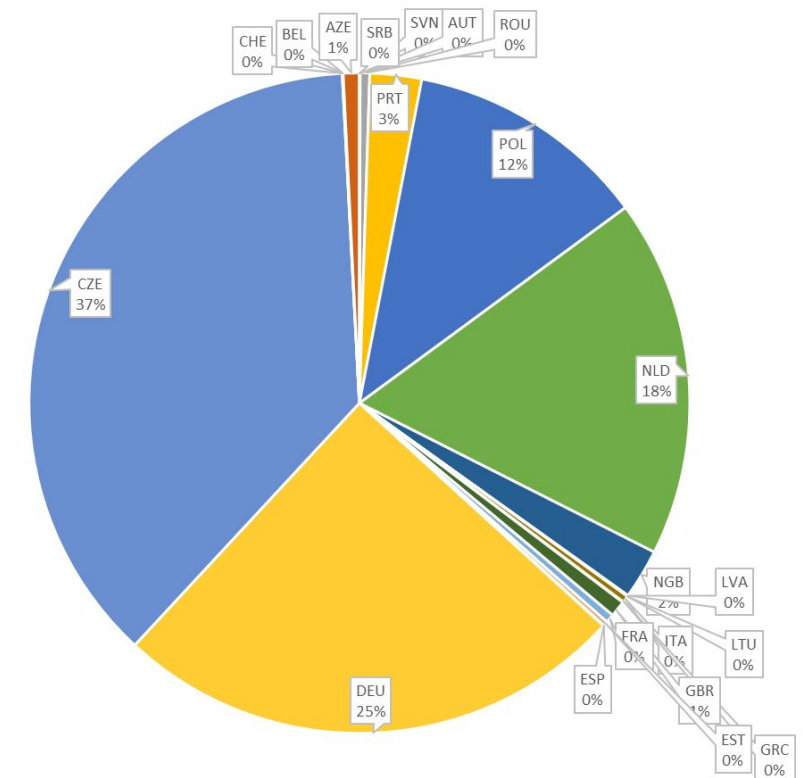
- **Optional metrics**

- **DOC_GEN (Number of accessions with accessible genomic information):**

- Can only be calculated by the genebanks themselves, if at all.

- **DOC_DOI (Number of accessions with a DOI):**

- 302,605 (14.3 %)



Conservation

Conservation

- Mandatory metrics
 - CON_NSAF (Number of accessions that are safety back-upped):
 - 921,258
- Optional metrics
 - CON_NDUP (Number of accessions that were safety duplicated to another genebank):
 - 822,401
 - CON_NTRI (Number of accessions that were safety duplicated to the Svalbard Global Seed Vault):
 - 119,445

Summary

- MCPD passport data only → no further management data
 - Thematic metrics groups
 - Conservation (except safety duplicates)
 - Availability
 - Distribution
 - Size and composition of the collection
 - Data and documentation
- Impossible to calculate with MCPD data
- Almost all metrics can be calculated
- Does it make sense to calculate the metrics in EURISCO/Genesys?
 - Calculation only possible for data that has been made available
 - Timeliness and scope of data determined by providers
 - Composition and documentation are fine, but no chance to give insight about conservation...
 - Alternatively, collect full metrics from the genebanks → Does this work?