UNIVERSITY OF HOHENHEIM

Core Collections

3502-470 Plant Genetic Resources

Prof. Karl Schmid
SS 2025

Institute of Plant Breeding, Seed Science and Population Genetics
University of Hohenheim

Overview

Background

How to establish a core collection

Examples of existing core collections

Background

How to establish a core collection

Examples of existing core collections

Core collections-necessity and concept

- Very large *ex situ* collections in gene banks
- They are too difficult to maintain and evalualate
- Sir Otto Frankel proposed use of core collections in 1984
- Definition: A subcollection of limited size that, while minimizes the similarities among its entries, represents the genetic diversity in the whole collection
- Core collections facilitate the use of PGR because they are smaller.

Characteristics of core collections

- Original definition: Limited set of accessions, minimum of repetitiveness, maximal resp. representitive diversity
- Within a given gene bank, a core collection represents the genetic diversity in the gene bank (core subset)
- Core collection for the whole species: synthetic core collection

Design parameters of core collections

- Contains <10% of whole collections
- Not more than 2000 entries

Why construct a core collection?

Goals

- Increase the efficiency of the evaluation and therefore the utilization of existing collections
- Produce a manageable and representative selection of germplasm for research and breeding
- Provide adequate material for standardisation in scientific work
- Rationalize the search for new traits (e.g., resistances)
- Accumulate large quantitities of diverse evaluation results for a limited set of accessions.
- Allele mining: efficiently find useful genetic variation

Methods for establishing core collections

- Random sampling: not so bad, since

> **Population genetic theory (Brown (1989))**
>
> On average, a random sample of 10% of accessions accounts for 70% of variation present in the total collection.

- Structured approach: Increases the captured variation while minimizing the number of accessions. For example, in 2000 the sorghum core collection from ICRISAT consisting of 3 % of accessions contained $> 90\%$ of the variation (see `http://www.icrisat.org/crop-sorghum-genebank.htm`)

Though more work, the structured approach is superior and should be used

---

Workflow of the structured approach

Step 1 Identify the material (collection) that will be represented by the core collection.

Step 2 Decide on the size of the core collection.

Step 3 Divide the set of material used into distinct groups.

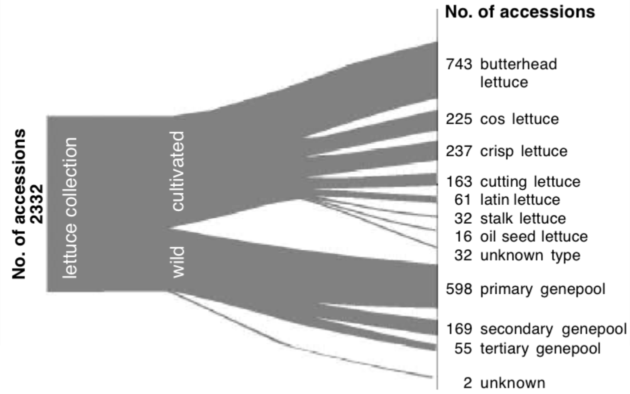
Step 4 Decide on the number of entries per group.

Step 5 Choose the entries from each group that will be included in the core.

Step 3: How to group material

- Geographic origin
- Phenotypic variation
- Genetic variation
- Agronomic characteristics (e.g., threshold for specific trait)

Example: Stratification of a Lactuca collection



Van Hintum et al. (2000)

Practical questions

Step 3: How to cluster into groups?

- Multivariate statistical approaches (e.g., PCA)
- STRUCTURE analysis

Step 4: How many individuals should be in each group?

- First idea: Same proportion of accessions in total collection should be in core collection (other ratios possible)

Make sure that all groups are incorporated. Small groups can contain important genetic information (e.g. about adaptation)

Using genetic markers to choose accessions

Advantages of genetic markers

- Relatively cheap
- Independent of the environment for classification
- Genetic distance and variation can be classified with good precision

Step 4/5: Strategies for choosing accesions

- H strategy ⇒ maximise heterozygosity
- M strategy ⇒ maximise allelic richness

The H strategy (Step 4/5)

- Estimate level of heterozygosity using a statistic like Nei's gene diversity h (estimated heterozygosity)
- Goal: assemble a collection in which gene diversity across markers is maximized
- Relative contribution of each group to total collection should be

$$h/(1-h)$$

  where h is the diversity within the group
- Statistical approach (Step 5): Pick individuals randomly within group

---

Step 4/5: The M strategy

- Focus on distinct allele types
- Maximize allelic richness
- Search for a group of accessions that maximizes the number of alleles in the core collection
- Number of accessions can be kept to a pre-defined number
- Deterministic approach (Step 5): The accessions to be included are selected by the algorithm

Step 5: Additional/alternative criteria for selecting accessions

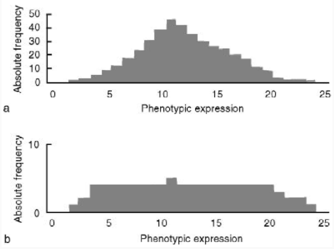
- Fulfill certain criteria (e.g., agronomic characteristics)
- Maximize diversity for all the other criteria
- Inclusion of external information (quality of data, politics, etc.)

Other methods

- Many other methods were developed (see James (2013))
- They differ in how to quantify diversity and how to choose a possible combination of accessions.

⇒ There is a nearly infinite number of possibilities to select 2000 accessions out of a total of 20000 accessions!

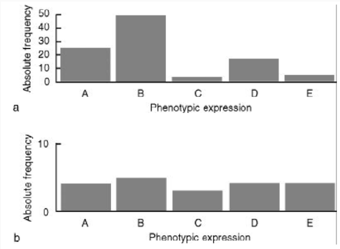
Distribution of variation in a core collection



Van Hintum et al. (2000)

a: in whole collection, b: in core collection

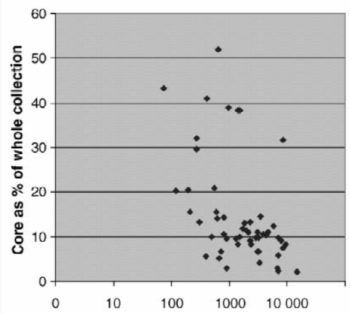
Distribution of variation in a core collection



Van Hintum et al. (2000)

a: in whole collection, b: in core collection

Relative sizes of core collections



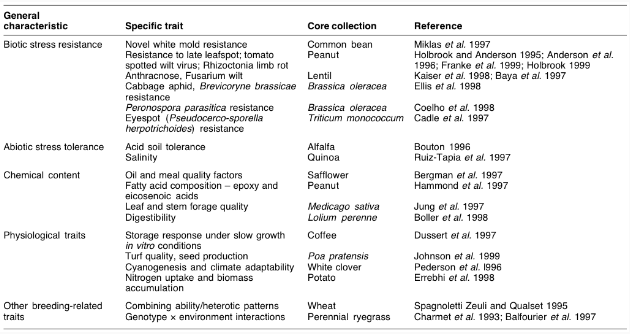
Van Hintum et al. (2000)

Background

How to establish a core collection

Examples of existing core collections

Several core collections used in the 1990s



General characteristic	Specific trait	Core collection	Reference
Biotic stress resistance	Novel white mold resistance	Common bean	Miklas *et al.* 1997
	Resistance to late leafspot; tomato spotted wilt virus; Rhizoctonia limb rot	Peanut	Holbrook and Anderson 1995; Anderson *et al.* 1996; Franke *et al.* 1999; Holbrook 1999
	Anthracnose, Fusarium wilt	Lentil	Kaiser *et al.* 1998; Baya *et al.* 1997
	Cabbage aphid, *Brevicoryne brassicae* resistance	*Brassica oleracea*	Ellis *et al.* 1998
	Peronospora parasitica resistance	*Brassica oleracea*	Coelho *et al.* 1998
	Eyespot (*Pseudocerco-sporella herpotrichoides*) resistance	*Triticum monococcum*	Cadle *et al.* 1997
Abiotic stress tolerance	Acid soil tolerance	Alfalfa	Bouton 1996
	Salinity	Quinoa	Ruiz-Tapia *et al.* 1997
Chemical content	Oil and meal quality factors	Safflower	Bergman *et al.* 1997
	Fatty acid composition – epoxy and eicosenoic acids	Peanut	Hammond *et al.* 1997
	Leaf and stem forage quality	*Medicago sativa*	Jung *et al.* 1997
	Digestibility	*Lolium perenne*	Boller *et al.* 1998
Physiological traits	Storage response under slow growth *in vitro* conditions	Coffee	Dussert *et al.* 1997
	Turf quality, seed production	*Poa pratensis*	Johnson *et al.* 1999
	Cyanogenesis and climate adaptability	White clover	Pederson *et al.* l996
	Nitrogen uptake and biomass accumulation	Potato	Errebhi *et al.* 1998
Other breeding-related traits	Combining ability/heterotic patterns	Wheat	Spagnoletti Zeuli and Qualset 1995
	Genotype × environment interactions	Perennial ryegrass	Charmet *et al.* 1993; Balfourier *et al.* 1997

Van Hintum et al. (2000)

Some recently established core collections (CC)

Crop	CC size	% collection	Reference
Foxtail millet (ICRISAT)	$\sim$ 155	10.5 %	Upadhyaya et al. (2008)
Sorghum landraces	107	3 %	Shezad et al. (2009)
Durum wheat (from Spain)	94	17 %	Carillo et al. (2013)
Mungbean (AVRDC)	289	4 %	Schafleitner et al. (2015)
Rice (USDA)	217	$\approx$ 1 %	Agrama et al. (2009)

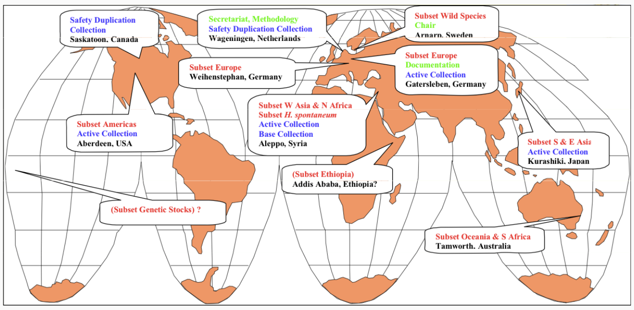
The International Barley Core Collection

- Established since 1989
- A synthetic collection
- Current size: $\approx$ 1300

Structure of the Barley Core Collection (2008)

Category	Count
Cultivars	500
Landraces	800
Wild barley *Hordeum vulgare* ssp. *spontaneum*	150-200
Other wild species (ca. 2 per species)	60 - 100
Genetic stocks, reference material	<200

Institutions involved in the Barley Core Collection



International Barley Core Collection

Summary

- Core collections are small subsets of large (ex situ) collections chosen to represent the variation in the whole collection
- Many strategies exist to choose the accessions for a core collection. Main idea: Identify groups of similar accessions, represent many groups with few accessions

---

Further reading

- Van Hintum et al. (2000), available via `http://www.bioversityinternational.org/nc/publications/publication/issue/core_collections_of_plant_genetic_resources.html`

Bibliography

Brown, A.H.D. (1989). The case for core collections. in: The Use of Plant Genetic Resources (A.H.D. Brown, O.H. Frankel, D.R. Marshall and J.T. Williams, eds.), p. 136–156. Cambridge University Press, Cambridge, UK

Jones, H., N. Gosman, R. Horsnell, G. A. Rose, L. A. Everest, A. R. Bentley, S. Tha, et al. (2013) Strategy for Exploiting Exotic Germplasm Using Genetic, Morphological, and Environmental Diversity: The *Aegilops tauschii* Coss. Example. Theoretical and Applied Genetics Online (5 April 2013). `doi:10.1007/s00122-013-2093-x`.

van Hintum, Th.J.L., Brown, A.H.D., Spillane, C., Hodgkin, T. (2000). Core collections of plant genetic resources. IPGRI Technical Bulletin 3

References i