Advances in Life Science and Technology
ISSN 2224-7181 (Paper) ISSN 2225-062X (Online)
Vol.101, 2025

Core Collection and Method of Developing Core Collection in Genetic Resource- A Review

Girma Ashe Ijara*
Ethiopia Institute of Agricultural Research, Addis Ababa, Ethiopia
P.O. Box 2003, Addis Ababa, Ethiopia

ABSTRACT
Agriculture today is characterized by a sharp reduction in the diversity of cultivated plant due to human and natural event. Plant breeding contributes to decrease of crop diversity through expansion of genetically homogeneous cultivars and promotion of few widely adapted varieties. The size of germplasm collections frequently restricts access to them and consequently, their use in plant breeding and research. Therefore management and use of germplasm collections could be enhanced if a limited number of genetically diverse accessions within the collection were selected as the core collection. The paper, therefore, aims to reviewing the way for core collection establishment and its implication in breeding program for crop improvement. A core collection is a subset of a large germplasm collection that involves of accessions selected to represent the genetic diversity of the collection. The core collection's objective is to improve the use and management of a germplasm collection. Creating a core collection is challenging and time taking process and can be done for any germplasm collection. Typically, accessions are grouped and selections are made from/within these groupings to create core collection. A basic process for creating a core collection can be divided into four steps, which include definition of domain, division in group, allocation of entries and choice of accession. A core collection provides a manageable sample size of the collection that is structured and smaller than the entire collection. Generally core collection is essential in crop improvement by simplifying the use of germplasm in gene bank operation, basic research and education.
Key words: Core collection, Germplasm, Germplasm Collection, Genetic resource, Genetic diversity.

DOI: 10.7176/ALST/101-03

Publication date: January 31st 2025

1. INTRODUCTION

1.2. Genetic Resource and Germplasm Collection

Agriculture today is characterized by a sharp reduction in the diversity of cultivated plants. Around 80,000 of the 350,000 plants species in the world are edible to humans. However, just around 150 species are being regularly cultivated, either for human food or as animal feed, and of these, 30 supply 95% of calories and protein consumed by humans. Four plants species rice (*Oryza sativa*), maize (*Zea mays*), wheat (*Triticum spp*.) and potatoes (*Solanum tuberosum*) are source of half world food (Fuleky, 2009). Plant breeding which includes the development of genetically homogenous cultivars and the promotion of a small number of widely adapted varieties, also reduce the intraspecific diversity of crop diversity in agriculture in addition to the interspecific reduction of crop diversity (Haussmann *et al*., 2004).

Genetic resources can be defined as all materials that are available for improvement of a cultivated plant species (Becker, 1993). In classical plant breeding, genetic resources may also be considered as those materials that, without selection for adaptation to the target environment, do not have any immediate use for the breeders (Hallauer and Miranda, 1981). According to the extended gene pool concept, genetic resources may be divided into primary gene pool, secondary gene pool, tertiary gene pool and isolated genes pool (Harlan and De Wet, 1971; Becker, 1993).

The primary gene pool consist of species itself and other species that can be easily crossed with it. The secondary gene consists of closely related species that are more challenging to cross with the intended crop, i.e. where cross-breeding progenies are partially sterile and cross-breeding is less effectives (low percentage of viable). The tertiary gene pools are species that can only be crossed through the use of specialized methods like embryo rescue or protoplast fusion. Isolated gene which makes up the fourth category of genetic resource can

come from animal or microbial source as well as from related or unrelated plants species. The significance of the different classes of genetic resources for crop improvement depends on the target crop species. In maize, for example, genetic variation in the primary gene pool is so large that the secondary or tertiary gene pools are rarely used. In rape seed, on the other hand, genetic variation in the primary gene pool is small and breeders have to transfer important traits from Brassica species of the secondary and tertiary gene pool into the cultivated species (Hu e*t al*., 2002).

Plant breeding is dependent on the germplasm, without which breeding is impossible. A species or population can be preserved by using its genetic material. Germplasm is useful for reproduction, but it can also be improved through genetic engineering or plant breeding to increase plant performance. The component (parents) needed to start a breeding program are provided by germplasm. Plant breeding occasionally limits themselves to assessing plant germplasm and selecting from already existing biological variation. Sometimes breeders create new variability by crossing parents, mutagenesis (causing mutations) and more recently, gene transfer. Following the application of suitable selection techniques to this base population, promising genotypes are identified and further assessed in preparation for releasing as cultivars. When plant breeders need to make improvement, they must located a source of germplasm that will provide the necessary genes for the breeding effort. Large amounts of germplasm must be collected, categorized, stored and management by certain organization (germplasm banks) in order to make use of it easier. This strategy gives researches easy access to germplasm whenever they need it (Acquaah, 2009).

In many cases the size of germplasm collections limits their accessibility, and thus their utilization in plant breeding and research. It can also limit the quality of their management. To improve this situation it has been proposed that a limited set of accessions be selected from a collection containing as much genetic diversity as possible. Such a selection would offer a good starting point when searching for new traits (Vaughan, 1991) and could be used for in-depth evaluation, thus increasing the knowledge of the entire collection (Knüpffer and van Hintum, 1995). Frankel (1984) introduced this concept, calling it a ‘’core collection’’.

Germplasm collections exist to conserve the genetic diversity of crop species and their wild relatives (Williams, 1991). Nevertheless, the size of many large germplasm collections may be an obstacle to their evaluation and utilization (Holden, 1984). The management and use of germplasm collections could be enhanced if a limited number of genetically diverse accessions within the collection were selected as the core collection (Frankel, 1984) [10] and given priority in evaluation and hybridization (Brown, 1989a). The papers, therefore aim to reviewing the way for core collection establishment and its implication in breeding program for crop improvement.

1.2. Core Collection

The collection and conservation of plant germplasm have made major progress during the previous several decades. Despite the fact that a lot of germplasm has been collected and their management has become more challenging due to their large sizes. Furthermore little is known about the interspecific and intraspecific genetic diversity and structure of such collection (Araceli *et al*., 2009). The idea of core collection has been put up to efficient use of large germplasm collection. A core collection is a subset of germplasm collection of a species that represent the genetic diversity of the entire collection (Frankel *et al*., 1984). According to Odong *et al*. (2013), a good core collection should not contain any redundant accessions, be manageable in size and represent the whole genetic diversity. A core collection offers a structured sample from the collection that is smaller than the entire collection and easier to manage. Its structure is such as to represent the diversity of the collection. It forma a reference set and when choices have to made, an automatic priority for attention (Van Hintum *et al*., 2000).

2. METHOD OF DEVELOPING CORE COLLECTION

Creating a core collection is time taking and challenging work. It can be work for every germplasm collection. Complete documentation or entirely trustworthy data are not necessary. Genetic marker information and advanced mathematical knowledge are not required. All that need is a collection of germplasm, occasionally some background information of the species being collected and the collection as whole, and some time to choose the core collection. However simple or complex a procedure is followed, it is always worth ensuring that there is consultation between gene bank managers, plant breeders and other research workers interested in the crop and the use of its genetic diversity (Van Hintum *et al*., 2000).

2.1. Criteria for good core collection

As Brown, (1989) suggest a good core collection must full fill the following criteria.
Utility criteria
1. The core should contain no redundant entries
2. The origins of its entries are authentic, unless no choice is available
3. It is sizeable enough to get accurate outcomes for the entire collection while still being an appropriate size.
4. It can forecast the source of useful information
Genetic criteria
1. The major sub-specific taxa and geographic regions are represented
2. Emphasis is given to representing the more broadly adapted rather than intensely specialized alleles
3. Within the constraints of criteria 1 and 2, genetic diversity, especially as measured by the number of alleles per locus, is maximal.

A core set can be selected based on a different data source, such as passport data, geographical origin (Thies *et al*., 2002, Quenouille *et al*., 2016) agronomic traits (Hanson *et al*., 2004, Zewdie *et al*., 2004, Fan *et al*., 2004) and molecular marker (Nicolaï *et al*., 2013). The goal of creating core collection is to maintain the diversity of the overall collection while reducing the number of representative accessions up to 10%. Depending on the research objective a number of possible approaches for creating core set are implemented. In the early 2000 years, most researchers performed random sampling using various methods (Thies *et al*., 2002, Hanson *et al*., 2004). Later, it was suggested that the M (maximum) strategy was more effective way to create a core set that represented the maximum genetic diversity without redundancy (Zewdie *et al*., 2004, Franco *et al*., 2005). But choosing a core collection that provides representative diversity is a more challenging. A general process for choosing a core collection can be divided into four stages, which will be briefly discussed in detail in the following sections (Van Hintum *et al*., 2000).

2.2. Definition of the domain

The first step in creating a core collection is defining the material that should be represented, in other words the domain of the core collection. The material that should be represented by the core, that is, the domain of the core collection will differ from core to core. It will depend on the material available and the purpose of the collection.

2.3. Division in groups

The division of the domain into distinct groups can be accomplished stepwise; this is, at every step a group of accessions is divided into as genetically distinct subgroups as possible. Van Hintum (1995) called this process branching, which can be done on the basis of taxonomy, domestication, distribution, breeding history and utilization, or on the basis of molecular marker or other characterization data. A 'diversity tree' can be constructed which represents the genetic structure of the domain. The exercise of branching ends when no further subdivision in genetically distinct groups can be made or sufficient branching was achieved.

2.4. Allocation of entries

The choice of the number of entries to be included in the core is arbitrary and will depend on the purpose of the core collection. Generally, the number of entries chosen is substantially lower than the number of accessions in the domain, usually in the 5 to 20% range.

The distribution of the entries among the group can also be done in a stepwise procedure. The way the entries in the group should be allocated among the subgroups is decided at each division of the group in to subgroups. In this procedure, there are various tactics to employ. The number of accessions in the subgroup can be utilized for this decision if information regarding the subgroups importance or diversity is not readily available. The constant (C), proportional (P) and logarithmic (L) techniques are accessible in this scenario (Brown 1989a). Quantitative information about the diversity in the subgroups is available in the gene diversity (H) or maximization (M) strategy can be used (Schoen and Brown 1995) and finally, if there is an idea of the relative importance of the subgroups this can also be used for the allocation.

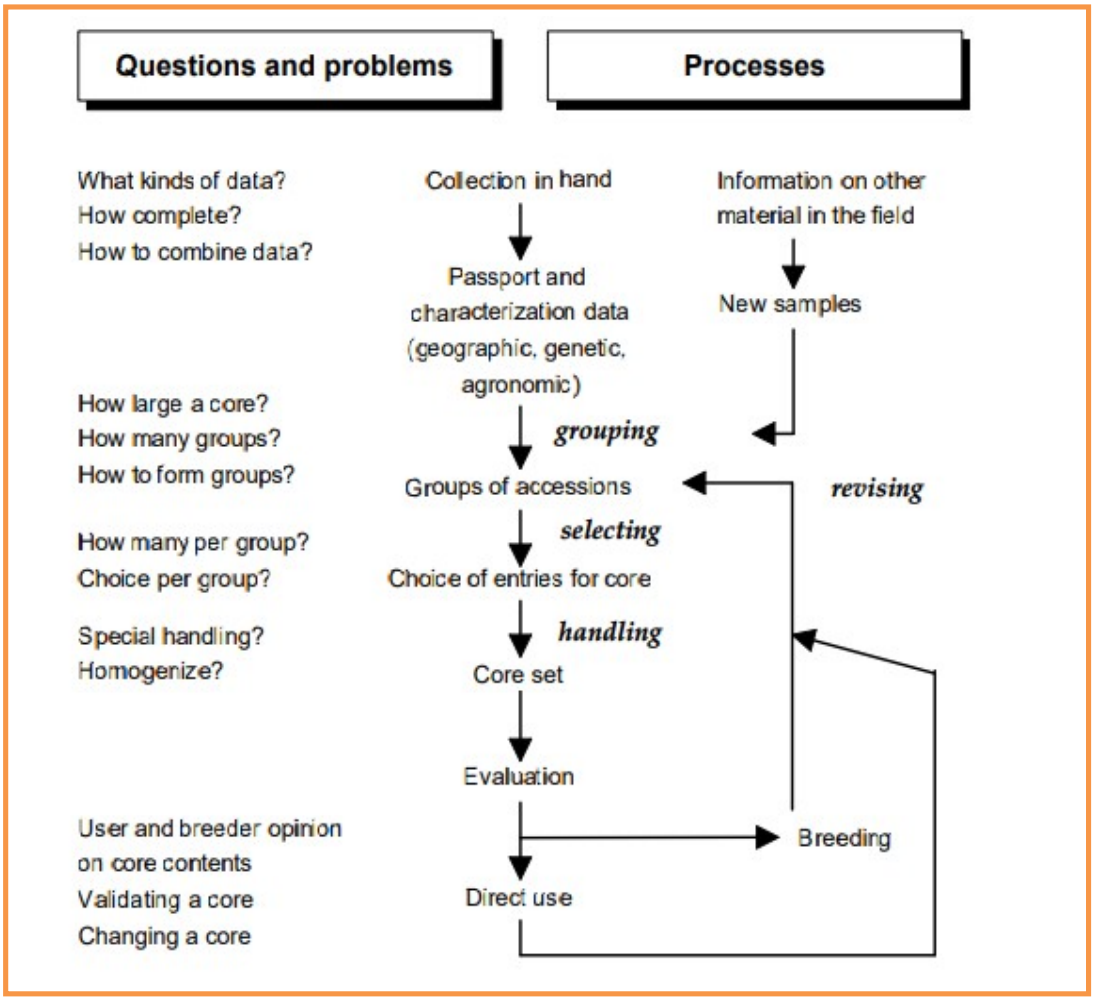
The method of distributing entries among subgroups will depend on the data that is available about the group and its subgroup, which can vary from group to group. For instance one may select the number of entries from each subgroup for the core collection based on the number of accession in the subgroups, if there is no information

available about the accession in a group of wild material that is divided into subgroups by country of origin. If the majority of the accessions in a group of cultivars have detailed molecular marker information available, one may use a diversity-based strategy. If the decision must be made regarding the division of cultivated and wild subgroups generally, a strategy based on the weight or importance of group for the core collection objective will be followed.

2.5. Choice of accessions

The final selection for particular entries from the grouping may be made at random or, if available on the basis of supplementary data. The goal should be to have the entries represent the group diversity as accurately as possible. Practical consideration may also be taken into account since it is preferable to have easily accessible, authentic and well documented material in the core set. This involves elements like accessibility of the seed and the accuracy and quantity of information regarding the accession.

If extensive and complete molecular marker data or other characterization data are available, it is possible to base the choice of entries on the result of a multivariate analysis. This would allow an optimal representation of diversity within the group.



Source: (Van Hintum *et al*., 2000)
Figure 1. Flow chart to illustrate steps in developing a core collection.

3. IMPLICATION OF CORE COLLECTION IN CROP IMPROVEMENT

A core collection provides a manageable sample from the collection that is organized and smaller than the entire collection. The collection diversity is reflected in its organization. When decision need to be made, it create a reference set and become an automatic priority

3.1. Improving Gene Bank Operations

The process of creating core collection typically result in a significant improvement in our understanding of the degree and distribution of genetic diversity within collection. Once it has been established, it serves as a starting point for further study. We expect that the core collection internal group and within group structures will exhibit patterns of diversity

Task	Function of core
Collection characteristics	
1 Dealing with new accessions	Provide a reference set for determining the group of accessions with which new material should be compared
2 Detection of gaps or uneven collecting	Allow identification of discontinuities in variation indicating missing material or sets where large amounts of variation are associated with a few accessions
Collection management	
3 Developing regeneration schedules	Core entries have priority, especially when upgrading collections
4 Prioritizing handling	Provide set for priority handling when needed
5 Monitoring viability	Provide appropriate set of accessions for monitoring whole
6 Duplication	Act as a priority group for safety-duplication, for further distribution to regional or international genebanks or for maintenance in different conditions (e.g. as DNA libraries, in field banks or *in vitro*)
7 Development and application of new conservation methods	Provide test material of choice for possible improved maintenance procedures (e.g. ultra-dry seeds, *in vitro* and cryopreservation)
Information management	
8 Database organization	Provide benchmark standard for documentation and allow stratification of whole collection to be recorded
Study and use of collections	
9 Developing descriptor lists	Entries appropriate to test sufficiency of descriptors to discriminate accessions
10 Testing expensive and complex traits (e.g. photoperiod response)	Allow development of efficient two-step sampling procedure – first between, then within groups
11 Method development	Provide set of material which is likely to cover full range of characteristic expression
12 Relationships between different characters	Provide restricted set likely to cover full range of different character expressions to maximize efficiency of correlation studies
13 Genetic studies	Allow selection of optimal material for studies of trait inheritance and estimation of general combining ability
14 Prebreeding	Provide dissimilar groups likely to assist in identifying heterosis or bringing together new gene combinations
Distribution from collections	
15 Having adequate seed supplies on hand	Larger amounts of seed can be produced of the limited set of core entries
16 Distribution of diverse and representative samples	Provide maximum diversity in limited sets of accessions for assessment by users

Source: (Van Hintum *et al*., 2000)
Figure 2. Examples of the ways in which core collections can be used

3.2. Improving the Use of Plant Genetic Resources

3.2.1. Screening for useful traits and characteristics

The core collection will frequently be used to screen for characteristics or traits that are advantageous for breeding purpose. For a wide range of qualitative and/or quantitative traits, core collection can be screened. The presence or absence of the traits or characteristics in the core might be used to make cost-benefit analyses of how comprehensive search strategies should be and what procedures will optimize benefits.

Core collections will be powerful tools where researchers are interested in background effects on gene expression (epistasis, epigenetics, etc.) and need a range of highly divergent background genotypes for the planned studies. Similarly, end users with an interest in assessing or exploiting the extent of either genotype by environment (G × E) interactions or gene or QTL by environment interactions can use core entries to assess the extent of such interactions in a gene pool (Charmet *et al*., 1993; Balfourier *et al*., 1997).

3.2.2. Basic research and education

Core collection will be helpful wherever a large a range of diversity is needed for either research or illustrative purposes. Crop gene pools size, genetic diversity distribution and the connection between them have all been studied using core collection (Tohme *et al*., 1996, 1999). In order to establish the association between cyanogenesis and climate in germplasm, core set may also be utilized to establish correlations between traits and environmental parameters (Pederson *et al*., 1996).

3.3. Increasing the Use of Core Collection.

Once a core collection has been created, it is crucial to inform users about it and how they could use it in their work. Users of germplasm shouldn't immediately be assumed to be aware of the presence and value of the core collection. Reaching as wide a range of users as possible with relevant information about the core collection will increase the utility of the collection to users and hence strengthen the link between the core and its users. The more customers that use the core and deliver response on its utility, the more the value and significance of the core to end user needs can be improved over time.

Future Direction

Establishing core collection is important in crop improvement as use of large number of genetic material in breeding purpose is time taking, cost and difficult in practical work. But germplasm material can be changed by the rapid use of biochemical and molecular process that result in new set of collection. On the other hand germplasm material can be exposed to changing environments and can thus adapt to new stress environment. So it is important to revising the original core collection from time to time.

4. REFERENCE

Acquaah, G., 2009. Principles of plant genetics and breeding. John Wiley & Sons

Araceli AM, Morrell PL, Roose ML, Kim SC. 2009. Genetic diversity and structure in semi-wild and domesticated chiles (*Capsicum annuum;* Solanaceae) from Mexico. *Am J Bot*. 96:1190–202.

Balfourier, F., J.A. Oliveira, G. Charmet and E. Arbones. 1997. Factorial regression analysis of genotype by environment interaction in ryegrass populations, using both isozyme and climatic data as covariates. *Euphytica* 98:37-47.

Becker, H., 1993. Pflanzenzu, chtung. Stuttgart: Verlag Eugen Ulmer. *Plant Genetic Resources*, *2*(1), pp.3-21.

Brown AHD 1989a. The case for core collections. In: Brown AHD, Frankel OH, Marshall DR, Williams JT (eds) The use of plant genetic resources. Cambridge University Press, Cambridge, pp 136-156

Brown, A.H.D., 1989. Core collections: a practical approach to genetic resources management. *Genome*, *31*(2), pp.818-824.

Charmet, G., F. Balfourier, C. Ravel and J.B. Denis. 1993. Genotype x environment interactions in a collection

of French perennial ryegrass populations. *Theory. Appl. Genet*. 86:731-736.

Fan M, Chen S, Engle LM. 2004. The study on genetic diversity of the core collection for Capsicum using random amplified. *J Agric Res China*. 53: 165–78.

Franco J, Crossa J, Taba S, Shands H. 2005. A sampling strategy for conserving genetic diversity when forming core subsets. *Crop Sci.* 45:1035–44.

Frankel, O.H. 1984. Genetic perspectives of germplasm conservation. Pp. 161-170 *in* Genetic Manipulation: Impact on Man and Society (W.K. Arber, K. Llimensee, W.J. Peacock and P. Starlinger, eds.). Cambridge University Press, Cambridge, UK.

Fuleky, G. ed. 2009. Cultivated Plants, Primarily as Food Sources, *Volume II*. EOLSS Publications.

Hallauer AR and Miranda JB. 1981. Quantitative Genetics in Maize Breeding. Ames: Iowa State University Press.

Hanson PM, Hanson PM, Yang R, Lin S, Tsou SCS, Lee TC. 2004. Variation for antioxidant activity and antioxidants in a subset of AVRDC-the World Vegetable Center Capsicum core collection. *Plant Genet Resour*. 2:153–66.

Harlan JR and de Wet JMJ. 1971. Toward a rational classification of cultivated plants. Taxon 20: 509–517.

Haussmann, B.I.G, Parzies, H.K, Presterl, T, Susic, Z. and Miedaner, T, 2004. Plant genetic resources in crop improvement. *Plant genetic resources*, *2*(1), pp.3-21.

Holden JHW. 1984. The second ten years. In: Holden JHW, Williams J (eds) Crop genetic resources: conservation and evaluation. George Allen and Unwin, London, UK, pp 277-285

Hu Q, Hansen LN, Laursen J, Dixelius C and Andersen SB. 2002. Intergeneric hybrids between Brassica napus and Orychophragmus violaceus containing traits of agronomic importance for oilseed rape breeding. Theoretical and Applied Genetics 105: 834–840.

Knüpffer, H. and Th.J.L. van Hintum. 1995. The barley core collection: an international effort. Pp. 171-178 *in* Core Collections of Plant Genetic Resources (T. Hodgkin, A.H.D. Brown, T.J.L. van Hintum and E.A.V. Morales, eds.). John Wiley and Sons, Chichester, UK.

Nicolaï M, Cantet M, Lefebvre V, Sage-Palloix AM, Palloix A. 2013. Genotyping a large collection of pepper (Capsicum spp.) with SSR loci brings new evidence for the wild origin of cultivated C. annuum and the structuring of genetic diversity by human selection of cultivar types. *Genet Resour Crop Evol.* 60:2375–90.

Odong TL, Jansen J, van Eeuwijk FA, van Hintum TJL. 2013. Quality of core collections for effective utilization of genetic resources review, discussion and interpretation. *Appl Genet.*126:289–305.

Pederson, G.A., Fairbrother, T.E. and Greene, S.L., 1996. Cynogenesis and climatic relationships in US white clover germplasm collection and core subset. *Crop Science*, *36*(2), pp.427-433.

Quenouille J, Saint-Felix L, Moury B, Palloix A. 2016. Diversity of genetic backgrounds modulating the durability of a major resistance gene. Analysis of a core collection of pepper landraces resistant to Potato virus Y. *Mol Plant Pathol*. 17:296–302.

Schoen, D.J. and Brown, A.H.D., 1995. Maximizing genetic diversity in core collections of wild relatives of crop species.

Thies JA, Fery RL. Evaluation of a core of the U.S. 2002. Capsicum germplasm collection for reaction to the Northern root-knot nematode. *Hort-Science.* 37:805–10.

Tohme, J., O.D. Gonzalez, S. Beebe and M.C. Duque. 1996. AFLP analysis of gene pools of a wild bean core

collection. Crop Sci. 36:1375-1384.

Tohme, J., S. Beebe and C. Iglesias. 1999. Molecular characterization of the CIAT bean and cassava core collections. Pp. 28-36 in Core Collections for Today and Tomorrow (R.C. Johnson and T. Hodgkin, eds.). IPGRI, Rome, Italy.

Van Hintum, Th.J.L. A.H.D. Brown, C. Spillane and T. Hodgkin. 2000. Core collections of plant genetic resources. IPGRI Technical Bulletin No. 3. International Plant Genetic Resources Institute, Rome, Italy.

Van Hintum, T.J., Von Bothmer, R. and Visser, D.L., 1995. Sampling strategies for composing a core collection of cultivated barley (*Hordeum vulgare*) collected in China. *Hereditas*, *122*(1), pp.7-17.

Vaughan, D.A. 1991. Choosing rice germplasm for evaluation. Euphytica 54:147-154.

Williams TJ .1991. Plant genetic resources: some new directions. *Advan Agron* 45:61-9 pp.

Zewdie Y, Tong N, Bosland P. 2004. Establishing a core collection of Capsicum using a cluster analysis with enlightened selection of accessions. *Genet Resour Crop Evol*. 51:147–51.