Core collection for enhance use and sustainable conservation of plant genetic resources: Case study using heuristic approach for Indian small millets collections

Prem Narain Mathur[1*], Yong-Jin Park[2], A. Seetharam[3] and Jayarame Gowda[3]

[1]South Asia Office, Bioversity International, NASC Complex, Pusa Campus, New Delhi 110012, India
[2]Department of Molecular Biotechnology, Konkuk University, 1 Hwayang-dong, Gwangjin-Gu, Seoul 143-701, Republic of Korea
[3]AICSMIP, University of Agricultural Sciences, GKVK Campus, Bangalore 560 065, Karnataka, India

Abstract

Increasing size of germplasm collections in most genebanks limits their accessibility for use in crop breeding and quality of their management. Therefore, it has been proposed that a limited set of accessions be selected containing as much genetic diversity as possible to offer a good starting point when searching for new traits, and could be used for in-depth evaluation, thus increasing the knowledge of the entire collection. Frankel (1984) introduced this concept, calling it a "core collection". Frankel and Brown (1984a and b) and Brown (1989a and b) further developed the concept. Many genebank curators around the world have now been working with core collection. As a result, the concept evolved from a theoretical idea into widely applied methods with many variations and has resulted in a wide variety of methodologies for identifying core entries. Basically, the procedure of creating a core collection is very simple and a random selection from a collection might be considered a core collection. But the selection of a core collection to ensure representative diversity is some what more complicated. Over the years, tremendous progress has been achieved using different methodologies, including stratified random sampling, and these methodologies have been successfully applied to develop core collections for various genepools. More recently, the Rural Development Administration (RDA) has used the advanced M strategy with a heuristic search for establishing core sets and accordingly a programme known as "PowerCore" has been developed and made available to users (http://genebank.rda.go.kr/powercore/). The programme support development of core-sets by reducing the redundancy of useful alleles and thus enhancing their richness. Output of the "PowerCore" has been validated using some case studies and the programme effectively simplifies the generation process of core-set while significantly cutting down the number of core entries, maintaining 100% of the diversity. "PowerCore" is applicable to various types of genomic data including single nucleotide polymorphisms (Kim *et al*. 2007). In the present study, "PowerCore" has been used to develop core sets of six Indian small millet collections viz., Barnyard millet (*Echinochloa frumentacea*), Finger millet (*Eleusine coracana*), Foxtail millet (*Setaria italica*), Kodo millet (*Paspalum scrobiculatum*), Little millet (*Panicum sumatrense*), and Proso or Common millet (*Panicum miliaceum*). The number of accession used for developing the core sets were: 729, 3924, 1478, 1038, 902 and 672 for Barnyard, Finger, Foxtail, Kodo, Little, and Proso millet, respectively. The validation of core sets using "PowerCore" has also been compared with the core developed using the stratified methodology. It has been observed that the core sets identified using "PowerCore" are small in size with greater diversity captured compared to traditional clustering methods used for all the six millets collections and the results has been discussed in this presentation.

Key Words

Core collection, PowerCore, Barnyard millet (*Echinochola frumentacea*), Finger millet (*Eleusine coracana*), Foxtail millet (*Setaria italica*), Kodo millet (*Paspalum scrobiculatum*), Little millet (*Panicum sumantrense*), Proso millet (*Panicum miliaceum*)

Introduction

In the context of limited resources available, 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 (Frankel 1984, Frankel and Brown 1984a and b and Brown 1989a and b). 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). Since the introduction of core collection concept, it has been interpreted and applied in many

different ways. This has resulted in a wide variety of methodologies for creating core collection and tremendous progress has been achieved using different methodologies and such methodologies have been successfully applied to develop core collections for various uses. However, the criteria by which core collections are established have intrinsic advantages and disadvantages for evaluating the genetic diversity of an entire collection. More recently, the Rural Development Administration (RDA) has used the advanced M strategy with a heuristic search for establishing core sets and accordingly a programme known as "PowerCore" has been developed which has been used for this study.

Materials and Methods

"PowerCore" was used for developing six small millet collections in India. The programme support development of core sets by reducing the redundancy of useful alleles and thus enhancing their richness. The programme is available for free download along with the user manual and sample data (http://genebank.rda.go.kr/powercore/). The six millet collections used for this study includes: Barnyard millet (*Echinochloa frumentacea*), Finger millet (*Eleusine coracana*), Foxtail millet (*Setaria italica*), Kodo millet (*Paspalum scrobiculatum*), Little millet (*Panicum sumatrense*), and Proso or Common millet (*Panicum miliaceum*). The number of accession used for developing core sets were: 729, 3924, 1478, 1038, 902 and 672 for Barnyard, Finger, Foxtail, Kodo, Little, and Proso millet, respectively. The numbers of quantitative and quality descriptors used were as per Bioversity descriptor lists and includes 24, 37, 25, 27, 21, and 19 descriptors for Barnyard, Finger, Foxtail, Kodo, Little, and Proso millet, respectively. Field characterization was undertaken during *Kharif* season (July-November) at Bangalore, India, which is situated at 13^0N latitude and $77^0$35' E longitude at an altitude of 890 meters. The average annual rainfall is around 850 mm, which is mostly received in the four rainfall months (July to October). The soil of the experimental field was red sandy loam with an acidic pH of 5.5. The soil are low in organic carbon (0.4%) with moderate availability of nitrogen (300kg/ha) and phosphorus (185kg/ha) and fairly rich in potash (225kg/ha). All accessions for each of the six millets, except Finger millet, were grown during the same year in an Augmented Block Design (Federer 1956) using three standard checks and hence the data for each of these five millets were used as single set for developing respective core collection. Since the total accessions were large for Finger millet, it was decided to characterise these accessions in four different years and accordingly 989, 989, 954 and 993 accessions were planted during *Kharif* season of 1987, 1988, 1989 and 1999, respectively also in an Augmented Block Design using three standard checks and the data were analysed for developing core for each year separately.

Results and Discussion

Core collections were considered to well represent the genetic diversity of the initial collection if the following two criteria were met: (1) no more than 20% of the traits had different means (significant at α=0.05) between the core collection and the initial collection and (2) Coincidence Rate (CR) was retained by the core collection in no less than 80% of the traits (Hu *et al.* 2000). The design concept and implementation strategy of "PowerCore" and the validation on the outcome in comparison with other methods has been well described by Kim *et. al.* (2007). "PowerCore" by default classified the continuous variables into different categories based on Sturges' rule (Sturges 1926), which is described as: $K = 1 + \log_2 n$, where n = number of observed accessions. However, the software also allows modifying this rule to make desired number of classes for the continuous variables. Once classification of the continuous variables is performed, the software takes into account all classes, without omission of any of its variables. Thus, possesses the capability to cover all the distribution ranges of each class.

For the present study maximum possible number of classes for each variable was used so that more diversity can be captured in core sets. The average per cent of core size for the nine sets was 7.30% with a range of 5.68 to 9.72% (Table 1). For the validation, the following statistical parameters were analysed to compare the mean and variance ratio between core and entire collections and are presented in Table 1: 1. **Mean Difference Percentage (MD %)** – which is estimated as: $MD\ (\%) = \frac{1}{m}\sum_{j=1}^{m}\frac{Me - Mc}{Mc} \times 100$, Where, Me = Mean of entire collection; Mc = Mean of core collection, and m = number of traits; **2. Variance Difference (VD %)** – which is estimated as: $VD\ (\%) = \frac{1}{m}\sum_{j=1}^{m}\frac{Ve - Vc}{Vc} \times 100$, Where, Ve = Variance of entire

collection, Vc = Variance of core collection, and m = number of traits; **3. Confidence Ratio (CR %)** – which is estimated as: CR (%) = $\frac{1}{m}\sum_{j=1}^{m}\frac{Rc}{\mathrm{Re}}\times 100$, Where, Re = Range of entire collection, Rc = Range of core collection, and m = number of traits. CR% indicates weather the distribution ranges of each variable in the core set are well represented when compared to the entire collection; and **4. Variable Rate (VR %)** – which is estimated as: VR (%) = $\frac{1}{m}\sum_{j=1}^{m}\frac{CVc}{CVe}\times 100$, Where, CVe = Coefficient of variation of entire collection, CVc = Coefficient of variation of core collection, and m = number of traits. VR% allows a comparison between the coefficient of variation values existing in the core collections and the entire collections and determines how well it is being represented in the core sets.

The results show that there was no significant difference (α=0.05) for the means of all traits between each of the nine core collections and the entire collections. CR% values for all core collections is greater than 80% and varied from 96.18 to 99.23% with an average value of 98.23%. Further, core collection with a larger VD% and VR% considered to provide a good representation of the genetic diversity of the entire collection. The average VD% for the nine core collections was recorded as 48.24% with a range of 43.11 to 53.25%, suggesting that variation observed in core collections was higher than the entire collection and the selected accessions were more dispersed and diverse. A higher value of VR% was also observed for all core collections with an average of 135.16% and a range of 126.41 to 145.83. The VR% greater than 100 for all the core collections indicated that the coefficient of variation for all traits in all nine core collections were greater in core collections compared to entire collections.

As discussed above, particular attention needs to be given to the high CR% while validating core collections compared to other statistical indicators used. Though "PowerCore" specifically indicates an exceptionally high CR% value for the core sets, 100% CR value is generally not obtained. The main reason for this is that in the case of continuous variables wherein classes are generated, "PowerCore" would only require the least number of accessions from each class. In view of the above, Kim *et al.* (2007) reported a new indicator, 'Coverage', which can be used to evaluate a core set for its coverage of variables. Coverage (%) = $\frac{1}{m}\sum_{j=1}^{m}\frac{Dc}{De}$ x 100, where Dc is the number of classes occupied in core collection and De is number of classes occupied in entire accessions in each character and m is the number of variables. The core sets resulted showed 100% coverage for all the nine core collections. This further suggests that "PowerCore" maintains all the diversity present in each class.

Table 1. Comparison of the percentage for the difference between the core collection and entire collections in millets

Core collections	Number of accessions			MD (%)	VD (%)	CR (%)	VR (%)	Coverage (%)
	Total	Core	(%)					
I-Barnyard millet	729	53	7.27	3.16	45.09	96.18	132.70	100
II-Finger millet:1987	988	71	7.19	4.44	49.98	98.84	136.36	100
III-Finger millet:1988	989	69	6.98	4.00	43.14	98.67	128.77	100
IV-Finger millet:1989	954	79	8.28	6.15	43.11	98.39	128.52	100
V-Finger millet:1990	993	77	7.75	6.72	49.36	98.31	134.6	100
VI-Foxtail millet	1478	88	5.95	3.71	49.08	99.23	145.83	100
VII-Kodo millet	1038	59	5.68	15.18	49.71	98.54	126.41	100
VIII-Little millet	895	87	9.72	8.06	53.25	98.00	138.89	100
IX-Proso millet	672	55	8.18	5.58	51.47	97.95	144.35	100

MD = Mean Difference, VD = variance Difference, CR = Coincidence rate, and VR = Variable Rate

Shannon-Wiener diversity index values were used for comparing the diversity pattern in core sets and the entire collections for each of the qualitative variables and it was observed that diversity index in most of the cases were higher for core accessions compared to entire collections, except for few variables, where it was observed at par.

Core collections were further validated by estimating mean and variance components for individual quantitative variable. It was observed that mean components (mean, minimum, maximum and rage) remains at par with the entire collections. High value for variance components for core collection compared to the entire collection for individual quantitative descriptors, indicating that core population is more diverse compared to the entire population even at the individual variable level.

The efficiency of "PowerCore" was also compared with clustering methods and it was observed that in most cases the mean component for the two core sets were at par and the variance components were higher in core collection developed using "PowerCore" compared to using clustering approach.

Conclusion

"PowerCore" is new faster approach for developing core collection, which effectively simplifies the generation process of a core set with a reduce number of core entries and maintaining high per cent of diversity compared to other methods used.

Acknowledgements

We are thankful to Dr. K.T. Krishne Gowda, Project Co-ordinator (Small Millets), Indian Council of Agricultural Research, for providing all necessary support for this study.

References

Brown, A.H.D. 1989a. Core collection: A practical approach to genetic resources management. Genome. 31:818-824.

Brown, A.D.H. 1989b. The case of core collections. Pp. 136-156 *in* The use of Plant Genetic Resources (A.H.D. Brown, O.H. Frankel, D.R. Marshall and J.T. Williams, eds.). Cambridge university Press, Cambridge, U.K.

Federer, W.T. 1956. Augmented (or hoonuiaku) designs. Hawaiian Planter's Record, v.55, p.191-208.

Frankel, O.H. 1984. Genetic perspective 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.

Frankel, O.H. and Brown, A.D.H. 1984a. Current plant genetic resources; A critical appraisal. Pp. 1-11 *in* Genetics: new frontiers, vol. 4. Oxford and IBH Publ. New Delhi, India.

Frankel, O.H. and Brown, A.D.H. 1984b. Plant genetic resources today: a critical appraisal. Pp. 249-257 in Crop Genetic Resources: Conservation and Evaluation (H.W. Williams and J.T. Williams, eds.). George Allen and Urwin, London.

Hu, J., Zhu, J. and Xu, H.M. 2000. Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. Theor. Appl. Genet. 101: 264-268.

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

Kyu-Won Kim, Hun-Ki Chung, Gyu-Taek Cho, Kyung-Ho Ma, Dorothy Chandrabalan, Jae-Gyun Gwag, Tae-San Kim, Eun-Gi Cho and Yong-Jin Park. 2007. PowerCore: A program applying the advanced M strategy with a heuristic search for establishing core mining sets. Bioinformatics 23(16): 2155-2162.

Sturges, H. 1926. The choice of a class-interval. J. Amer. Statist. Assoc. 21: 65-66.

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