Check for updates

Received: 4 April 2022 | Revised: 4 July 2022 | Accepted: 17 August 2022

DOI: 10.1002/leg3.158

ORIGINAL ARTICLE

Legume Science Open Access WILEY

Development of core collections in soybean on the basis of seed size

Ramakrishnan M. Nair[1] | Miao-rong Yan[2] | Anil Kumar Vemula[3] | Abhishek Rathore[3] | Maarten van Zonneveld[2] | Roland Schafleitner[2]

[1]World Vegetable Center South Asia, ICRISAT Campus, Patancheru 502 324, Hyderabad, Telangana, India

[2]World Vegetable Center, PO Box 42, Shanhua, Tainan, 74199, Taiwan

[3]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), 502 324 Greater, Hyderabad, Telangana, India

Correspondence
Ramakrishnan M. Nair, World Vegetable Center South Asia, ICRISAT Campus, Patancheru 502 324, Hyderabad, Telangana, India.
Email: ramakrishnan.nair@worldveg.org

Funding information
WorldVeg Core Funding

Abstract

Core collections display a large fraction of the diversity contained in large collections in smaller germplasm panels. We used historical data (1973–2015) collected at the World Vegetable Center, Taiwan, for developing soybean (*Glycine max* L. Merr.) core collections representing the diversity of the whole collection of 7853 accessions held by the Center. The collection was split into two groups on the basis of the 100 seed weight: large seeded (>25 g or equal to 25 g) and small seeded (<25 g). The large-seeded group (vegetable soybean/edamame) comprised 456 accessions, while the small-seeded group contained 7397 accessions. Within these two groups, we developed core collections based on seven quantitative and 14 qualitative traits collected during the autumn season, resulting in a core collection of 112 large-seeded vegetable soybean accessions and 1480 accessions for the small-seeded types.

KEYWORDS
core collection, grain soybean, vegetable soybean

1 | INTRODUCTION

Soybean, *Glycine max* (L.) Merr., is an important protein and oilseed crop. It serves as a high-quality protein source for livestock and aquaculture and a source of oil and other compounds for industrial uses and is a valued component of human diets. Vegetable soybean, also known as *edamame*, is a specialty type of soybean harvested near the full-seed stage (Zhang et al., 2015). In contrast to grain-type soybean, vegetable soybean cultivars have been selected for large seeds, a characteristic flavour called 'umami' in Japan (Torii, 1987), and bright-green well-shaped pods (Masuda, 1991).

Soybean crop improvement depends on access to the biodiversity of the species, which is the source of variation that sustains the genetic gain in breeding programmes. Resistances to pests and diseases and tolerance to abiotic stresses absent in cultivars as well as traits to adapt to new production methods or respond to new market demands can be sourced from biodiverse germplasm. Genebanks conserve the global diversity of soybean, a diversity that is at risk to be lost, as genetically improved high-yielding and uniform varieties replace more biodiverse landraces. Genebanks also conserve crop wild relatives of soybean, which may be at risk of extinction due altered land use and climate change.

The Genesys database of plant genetic resources for food and agriculture lists 266,252 *Glycine* accessions that are conserved ex situ in genebanks worldwide, with the largest collections held by the Australian Pastures Genebank (84,995 accessions), the International Centre for Agricultural Research in Dry Areas, Lebanon (27,014 accessions), Centro Internacional de Agricultura Tropical (22,694), USDA-ARS (22,490) and Embrapa, Brazil (21,637 accessions). In addition, large collections not listed in Genesys are held by the National Genebank of China (over 30,000 unique accessions of soybean; Chang et al., 1996; Chang & Sun, 1991; Wang, 1982) and by the National Agrobiodiversity Center of Rural Development Administration (approximately 20,000 accessions; http://genebank.rda.go.kr/). The

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2022 The Authors. *Legume Science* published by Wiley Periodicals LLC.

World Vegetable Center (WorldVeg) holds a *Glycine* spp. collection of 13,445 accessions. These collections include 17 species, of which there are 12,258 accessions of *G. max*. Identifying new traits of interest in large germplasm collections is laborious and costly. Establishing subset of collections, either core collections, which represent the diversity of the whole collection, or subsets enriched for specific traits makes screening more practical (El Bouhssini et al., 2011; van Hintum, 1999). The smaller size of core collections facilitates sharing of the germplasm with collaborators and saves labour and costs for germplasm screening. Various soybean core and mini-core collections have been established, for example, at USDA in the United States (Oliveira et al., 2010), in China (Song et al., 2010; Wang et al., 2006), in Korea (Jeong et al., 2019), Brazil (Priolli et al., 2013) and Japan (Kuroda et al., 2009). A core collection for Taiwanese vegetable soybean has been described by Kao et al. (2021).

The objectives of this research were to characterize the *G. max* accessions of the WorldVeg genebank using available passport data and standard morphological descriptors and analyse the diversity of the collection and to define a core subset as a valuable entry point for further research or use of the collection by plant breeders.

2 | MATERIALS AND METHODS

A collection of 7853 soybean accessions originating from all continents except Antarctica plus 75 accessions lacking information of origin was utilised to generate core collections. Evaluation data on seven quantitative and 14 qualitative traits were obtained from 1973 to 2015. The datasets included number of seeds per pod, plant height at R1 stage, plant height at R8 stage, number of primary branches, days to 50% flowering, number of pods per plant and 100-seed weight under quantitative and Pubescence colour, Pubescence type, Hypocotyl colour, Corolla colour, Mature pod colour, Seed colour, Seed coat pattern, Hilum colour, Seed coat surface lustre, Stem determination, Leaflet shape, Pubescence density, Leaflet size and Lodging score under qualitative traits.

The whole collection was divided into large-seeded (> = 25 g) (456 accessions) and small-seeded (<25 g) (7397 accessions) groups by setting a threshold on 100 seed weight obtained in the autumn harvest (100-seed weight) (Table S1). The rationale for this threshold was to distinguish the vegetable soybean types (larger seeded) from the grain types.

Data on 21 quantitative and qualitative traits were standardized using the range of each trait to eliminate the scale differences (Milligan & Cooper, 1985). This standardized data were subjected to hierarchical clustering using the DGOWER distance method (Gower & Legendre, 1986) at an R^2 (squared multiple correlation) value of 0.75 using SAS (SAS Institute Inc, 2018), and then by simple random sampling without replacement, 20% of the accessions (core subset) were drawn from each cluster.

In order to test whether the core subset represents the whole collection data, comparison of the means and variance between the whole and core collection were done by *t*-test (Snedecor & Cochran, 1989) and *F*-test, respectively. The distribution homogeneity for each of the seven quantitative traits among the whole collection and core subset was analysed by χ^2 test. Wilcoxon rank-sum non-parametric test was performed for the seven quantitative traits with the SAS NPAR1WAY procedure (SAS Institute Inc, 2018) to determine whether the whole collection represents the core subset for each of these traits. The phenotypic correlation among different traits was estimated independently in the whole collection and the core subset using Karl Pearson's correlations. The proportion of variance in one trait that can be attributed to its linear relationship with a second trait is indicated by the square of the correlation coefficient (Snedecor & Cochran, 1980). Based on this criterion, the correlation coefficients with an absolute value greater than 0.71 have been suggested to indicate correlation (Skinner et al., 1999) predicting more than 50% of the trait variation by the variation of another trait. Shannon–Weaver diversity index (Shannon & Weaver, 1949) was calculated for the whole collection and core subset data to determine to which extent the diversity for each trait was retained in core subset.

For an individual trait, a low value of H` (<0.50) indicates an unstable frequency class and lack of genetic diversity. Average H` of 0.50 < H` < =0.75 indicates the genetic diversity was moderate, and H` > 0.75 indicates high genetic diversity. All statistical analysis was done in SAS v9.4 (SAS Institute Inc., 2018).

3 | RESULTS

3.1 | Random selection of accessions from each cluster

From the large-seeded group, out of 456 accessions, a 20% core subset of 112 accessions was randomly drawn from each cluster consisting of 31 accessions from Asia, 48 from North America and eight of unknown origin. Due to a smaller number of accessions from Africa, Australia, Europe and South America, accessions from these regions were not part of the randomly drawn set. In the small-seeded group, out of 7397 accessions, the core collection of 20% (1480 accessions) contained 77 entries from Africa, 701 from Asia, 80 from Australia, 140 from Europe, 444 from North America, 29 from South America and nine of unknown origin.

3.2 | Large-seeded group

The differences for means and variances between whole collection and core subset were non-significant (Prob> = 0.05), and the distribution of trait values was homogeneous for all seven quantitative traits (Table 1). These results demonstrated that the core subset is representative of the whole collection and that the variation contained in the large was preserved in the core. The analysis of frequency distribution indicated homogeneous distribution of the trait values of all descriptors among whole collection and core subset except for plant height at R8 stage (Table 2). Karl Pearson's correlations of among

26396181, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/leg3.158, Wiley Online Library on [08/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Mean and variances for seven quantitative traits in whole and core subset of soybean for large- and small-seeded group

	Large-seeded group						Small-seeded group					
	Mean			Variance			Mean			Variance		
Trait	Whole	Core	Prob	Whole	Core	Prob	Whole	Core	Prob	Whole	Core	Prob
SN	1.42	1.43	0.40	0.09	0.10	0.26	1.52	1.52	0.39	0.09	0.09	0.41
PH_R1	21.60	22.57	0.26	8.22	7.59	0.31	29.32	29.28	0.93	14.78	14.70	0.80
PH_R8	23.80	25.95	0.06	10.55	11.47	0.25	39.12	39.12	0.99	20.96	21.03	0.86
NP	1.52	1.54	0.62	0.30	0.33	0.14	1.46	1.47	0.62	0.34	0.35	0.04
DF	5.52	5.47	0.28	0.42	0.43	0.70	5.61	5.61	0.89	0.56	0.58	0.08
PN	4.20	4.36	0.17	1.02	1.16	0.07	4.87	4.88	0.85	1.42	1.43	0.77
SW	29.42	29.43	0.98	4.27	4.32	0.85	14.65	14.63	0.84	4.49	4.59	0.26

Abbreviations: DF, days to flowering; NP, number of primary branches; PH_R1, plant height at R1 stage; PH_R8, plant height at R8 stage; PN, number of pods per plant; SN, number of seeds per pod; SW, 100-seed weight.

TABLE 2 Chi-square and probability for comparisons of frequency distribution for seven quantitative traits in both whole and core subset of soybean for large- and small-seeded group

	Large-seeded group		Small-seeded group	
Trait	χ^2	Prob	χ^2	Prob
SN	0.651	0.42	0.471	0.49
PH_R1	2.512	0.11	0.008	0.93
PH_R8	4.134	0.04	0.006	0.94
NP	0.342	0.56	0.036	0.85
DF	1.253	0.26	0.039	0.84
PN	0.412	0.52	0.076	0.78
SW	0.045	0.83	0.001	0.99

Abbreviations: DF, days to flowering; NP, number of primary branches; PH_R1, plant height at R1 stage; PH_R8, plant height at R8 stage; PN, number of pods per plant; SN, number of seeds per pod; SW, 100-seed weight.

phenotypic values was conducted for seven quantitative traits for the whole collection and core subset independently. Accordingly, a meaningful relationship was identified between plant height at R1 stage and R8 stage for both whole and core subset: r = 0.88 and 0.82, respectively. This core subset preserved the phenotypic correlations observed in the whole collection (Table 3). The 100-seed weight was significantly associated with number of seeds per pod (r = −0.30) and number of primary branches (r = 0.19), and similarly, number of pods per plant was significantly positively associated with number of seeds per pod (r = 0.20), plant height at R1 and R8 stages (r = 0.26 and 0.32, respectively) and number of primary branches (r = 0.66) in the core subset, suggesting that 100-seed weight and number of pods per plant could serve a useful purpose in choosing the accessions for higher yield.

Shannon–Weaver diversity index (H`) was calculated to compare phenotypic diversity for each trait (Table 4). Among seven quantitative traits, all traits showed high genetic diversity, except 100-seed weight for both whole collection and core subset. Among 14 qualitative traits, stem determination (SD) and leaflet shape (LLS) indicated poor genetic diversity. Hypocotyl colour (HCC), corolla colour (CC), mature pod colour (MPC) and leaflet size (LS) indicated moderate genetic diversity in both whole and core subset. From Table 4, it is clear that the diversity of the whole collection was represented in the core subset, with the range of 87%–99% for quantitative traits and 84%–99% for qualitative traits except LLS and SD.

3.3 | Small-seeded group

Like for the large-seeded group, the differences of means and homogeneity for all seven quantitative traits remained insignificant between whole collection and core subset (Prob> = 0.05), and homogeneity of variances between whole and core subset was non-significant for all quantitative traits except number of primary branches (Prob < 0.04) (Table 1). Also here the results indicated that the core subset is representative of the whole collection and that the variation was preserved. The analysis of frequency distribution indicated that all descriptors showed homogeneous distribution in the whole collection and the core subset (Table 2). Karl Pearson's correlations among phenotypic traits indicated meaningful relationship between plant height at R1 stage and R8 stage for both whole and core subset at r = 0.84, and overall this core subset, like for the large-seeded accessions, preserved the phenotypic correlations observed in the whole collection (Table 3). Consequently, 100-seed weight is significantly and negatively correlated with remaining all six traits, and similarly, number of pods per plant was significantly and positively correlated with number of seeds per pod (r = 0.36), plant height at R1 and R8 stages (r = 0.17 and 0.42, respectively), number of primary branches (r = 0.46) and days to flowering (r = 0.31) and negatively correlated with 100 seed weight (r = −0.18) in core subset, suggesting that 100-seed weight and number of pods per plant could serve a useful purpose in choosing the accessions for higher pod yield.

26396181, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/leg3.158, Wiley Online Library on [08/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Karl Pearson's correlation coefficients between seven traits in whole collection and core subset of soybean for large- and small-seeded group

		SN	PH_R1	PH_R8	NP	DF	PN
Large-seeded group	PH_R1	0.02 (−0.03)					
	PH_R8	0.04 (−0.16)	0.88** (0.82**)				
	NP	−0.02 (−0.12)	0.20** (−0.17)	0.16** (−0.15)			
	DF	0.04 (−0.11)	−0.42** (−0.45**)	−0.43** (−0.32**)	−0.20** (−0.10)		
	PN	0.26** (0.20*)	0.31** (0.26**)	0.36** (0.32**)	0.48** (0.66**)	−0.17** (−0.001)	
	SW	−0.24** (−0.30**)	0.22** (−0.10)	0.18** (−0.03)	0.22** (0.19*)	−0.14** (−0.15)	0.03 (−0.03)
Small-seeded group	PH_R1	0.27** (0.23**)					
	PH_R8	0.38** (0.34**)	0.84** (0.84**)				
	NP	0.16** (0.14**)	0.16** (0.16**)	0.21** (0.22**)			
	DF	−0.10** (−0.12**)	0.27** (0.26**)	0.35** (0.34**)	0.15** (0.16**)		
	PN	0.40** (0.36*)	0.14** (0.17**)	0.38** (0.42**)	0.45** (0.46**)	0.30** (0.31**)	
	SW	−0.25** (−0.25**)	−0.47** (−0.45**)	−0.49** (−0.49**)	−0.17** (−0.17**)	−0.18** (−0.16**)	−0.17** (−0.18**)

Note: In brackets are correlation coefficient for core subset.
Abbreviations: DF, days to flowering; NP, number of primary branches; PH_R1, plant height at R1 stage; PH_R8, plant height at R8 stage; PN, number of pods per plant; SN, number of seeds per pod; SW,100-seed weight.

Shannon–Weaver diversity index (H`) was calculated to compare phenotypic diversity among characters for each trait (Table 4). Among the seven quantitative traits, all traits showed high genetic diversity with range from 0.74 to 0.93 and 0.74 to 0.89 for the whole and core subset, respectively. Among the 14 qualitative traits, seed coat pattern (SCP) and leaflet shape (LLS) showed poor genetic diversity for both the whole collection and the core subsets. LS indicated moderate genetic diversity in both whole and core subset and the remaining 11 traits showed similar high genetic diversity in the WC and CC (Table 4), clearly indicating that the diversity of the whole collection was represented in the core subset with the range of 94%–100% and 95%–100% for both quantitative and qualitative traits respectively.

The core subsets for large-seeded (112) and small-seeded group (1480), which represents 20% of the whole collection (large seeded—456 and small seeded—7397) should represent the total diversity contained in the whole collection.

TABLE 4 Shannon–Weaver phenotypic diversity index of quantitative and qualitative traits for whole collection and core subset of soybean for large- and small-seeded group

		Large-seeded group		Small-seeded group	
Category	Trait	Whole	Core	Whole	Core
Quantitative	SN	0.80	0.76	0.93	0.87
	PH_R1	0.78	0.90	0.82	0.83
	PH_R8	0.76	0.75	0.83	0.84
	NP	0.78	0.85	0.82	0.78
	DF	0.98	0.93	0.75	0.79
	PN	0.84	0.85	0.74	0.74
	SW	0.61	0.62	0.88	0.89
Qualitative	PC	1.07	1.08	1.00	1.00
	PT	0.86	0.77	0.87	0.85
	HCC	0.61	0.59	0.62	0.62
	CC	0.62	0.6	0.61	0.61
	MPC	0.56	0.55	0.68	0.69
	SC	1.40	1.24	1.07	1.07
	SCP	0.89	0.83	0.42	0.44
	HC	1.51	1.55	1.27	1.28
	SCSL	0.91	0.93	0.99	0.99
	SD	0.19	0.27	0.84	0.85
	LLS	0.10	0.18	0.39	0.40
	PD	1.13	1.11	1.11	1.09
	LS	0.63	0.69	0.52	0.52
	LDS	0.82	0.98	1.38	1.39

Abbreviations: CC, Corolla colour; DF, days to flowering; HC, Hilum colour; HCC, Hypocotyl colour; LDS, Lodging score; LLS, Leaflet shape; LS, Leaflet size; MPC, Mature pod colour; NP, number of primary branches; PC, Pubescence colour; PD, Pubescence density; PH_R1, plant height at R1 stage; PH_R8, plant height at R8 stage; PN, number of pods per plant; PT, Pubescence type; SC, Seed colour; SCP, Seed coat pattern; SCSL, Seed coat surface lustre; SD, Stem determination; SN, number of seeds per pod.

26396181, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/leg3.158, Wiley Online Library on [08/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4 | DISCUSSION

An adequate and proper sampling is essential in developing a representative core collection and should consider the conservation of phenotypic associations arising out of co-adapted gene complexes (Ortiz et al., 1998). The analyses performed indicated that the selection of the core subset for both large- and small-seeded accessions was adequate and that the co-adapted gene complexes controlling these associations were sampled properly and adequately. Other relationships that did not meet the 50% criterion (Skinner et al., 1999) may be of interest to breeders.

Our vegetable soybean core collection complements another recently developed vegetable soybean core collection (Kao et al., 2021). Our collection is smaller and has a broader geographic range compared to the latter collection. In comparison to other soybean core collections, the WorldVeg collection contains a larger proportion of large-seeded vegetable soybean accessions.

The definition of this core collection was done on the basis of a unique morphological characterization dataset that was collected between 1973 and 2015. The collection has not yet been genotyped. That type of data can provide further information as shown by Kao et al. (2021).

The distinction between large- and small-seed soybeans will help to improve seed storage of soybean because large-seeded soybean accessions have usually a lower seed longevity (Wien & Kueneman, 1981). Distinct regeneration and seed monitoring protocols can be developed for these two types of soybean. We expect that large-seed soybean require more frequent seed monitoring and regeneration compared to small-seeded soybean types.

The definition of these core collections will help to enhance the use and management of the WorldVeg collection of soybean. The core collections are being published in Genesys and can be requested for research and breeding under the Standard Material Transfer Agreement (SMTA).

ACKNOWLEDGMENTS

The support of the long-term strategic donors of the World Vegetable Center, namely, Taiwan, United States Agency for International Development (USAID), UK Government's Foreign, Commonwealth & Development Office (FCDO), Australian Centre for International Agricultural Research (ACIAR), Germany, Thailand, Philippines, Korea and Japan is acknowledged.

CONFLICT OF INTEREST

All authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Ramakrishnan M. Nair https://orcid.org/0000-0002-2787-8396

REFERENCES

Chang, R. Z., & Sun, J. Y. (1991). *Catalogues of Chinese soybean germplasm and resources: Continuation I*. China Agricultural Press. (in Chinese)

Chang, R. Z., Sun, J. Y., Qiu, L. J., & Chen, Y. W. (1996). *Catalogues of Chinese soybean germplasm and resources: Continuation II*. China Agricultural Press. (in Chinese)

El Bouhssini, M., Street, K., Amri, A., Mackay, M., Ogbonnaya, F. C., Omran, A., Abdalla, O., Baum, M., Dabbous, A., & Rihawi, F. (2011). Sources of resistance in bread wheat to Russian wheat aphid (*Diuraphis noxia*) in Syria identified using the focused identification of germplasm strategy (FIGS). *Plant Breeding*, *130*, 96–97. https://doi.org/10.1111/j.1439-0523.2010.01814.x

Gower, J. C., & Legendre, P. (1986). Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification*, *3*, 5–48. https://doi.org/10.1007/BF01896809

Jeong, N., Kim, K. S., Jeong, S., Kim, J. Y., Park, S. K., Lee, J. S., Jeong, S. C., Kang, S. T., Ha, B. K., Kim, D. Y., & Kim, N. (2019). Korean soybean core collection: Genotypic and phenotypic diversity population structure and genome-wide association study. *PLoS One*, *14*(10), e0224074.

Kao, C. F., He, S. S., Wang, C. S., Lai, Z. Y., Lin, D. G., & Chen, S. (2021). A modified Roger's distance algorithm for mixed quantitative–qualitative phenotypes to establish a core collection for Taiwanese vegetable soybeans. *Frontiers in Plant Science*, *11*, 2143. https://doi.org/10.3389/fpls.2020.612106

Kuroda, Y., Tomooka, N., Kaga, A., Wanigadeva, S. M., & Vaughan, D. A. (2009). Genetic diversity of wild soybean (*Glycine soja* Sieb. et Zucc.) and Japanese cultivated soybeans [*G. max* (L.) Merr.] based on microsatellite (SSR) analysis and the selection of a core collection. *Genetic Resources and Crop Evolution*, *56*(8), 1045–1055.

Masuda, R. (1991). Quality requirement and improvement of vegetable soybean. In S. Shanmugasundaram (Ed.), *Vegetable soybean: Research needs for production and quality improvement* (pp. 92–102). Asian Vegetable Res. Dev. Center.

Milligan, G. W., & Cooper, M. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, *50-2*, 159–179.

Oliveira, M. F., Nelson, R. L., Geraldi, I. O., Cruz, C. D., & de Toledo, J. F. (2010). Establishing a soybean germplasm core collection. *Field Crops Research*, *119*(2–3), 277–289.

Ortiz, R., Ruiz-Tapia, E. N., & Mujica-Sanchez, A. (1998). Sampling strategy for a core collection of Peruvian quinoa germplasm. *Theoretical and Applied Genetics*, *96*, 475–483. https://doi.org/10.1007/s001220050764

Priolli, R. H., Wysmierski, P. T., Cunha, C. P., Pinheiro, J. B., & Vello, N. A. (2013). Genetic structure and a selected core set of Brazilian soybean cultivars. *Genetics and Molecular Biology*, *36*, 382–390. https://doi.org/10.1590/S1415-47572013005000034

SAS Institute Inc. (2018). SAS/STAT® 15.1 user's guide. Cary, NC: SAS Institute Inc.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. University of Illinois Press.

Skinner, D. Z., Bauchan, G. R., Auricht, G., & Hughes, S. (1999). A method for the efficient management and utilization of large germplasm collections. *Crop Science*, *39*, 1237–1242. https://doi.org/10.2135/cropsci1999.0011183X003900040046x

Snedecor, G. W., & Cochran, W. G. (1980). *Statistical methods* (7th ed.). Iowa State University Press.

Snedecor, G. W., & Cochran, W. G. (1989). *Statistical methods* (Eighth ed.). Iowa State University Press.

Song, X., Li, Y., Chang, R., Guo, P., & Qiu, L. (2010). Population structure and genetic diversity of mini core collection of cultivated soybean (*Glycine max* [L] Merr.) in China. *Scientia Agricultura Sinica.*, *43*(11), 2209–2219.

Torii, K. (1987). In Y. Kawamura & M. R. Kare (Eds.), *Umami; a basic taste*. Marcel Dekker.

26396181, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/leg3.158, Wiley Online Library on [08/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

van Hintum, T. J. L. (1999). The Core selector, a system to generate representative selections of germplasm accessions. *Plant Genetic Resources Newsletter.*, *118*, 64–67.

Wang, G. X. (1982). *Catalogues of Chinese soybean germplasm and resources*. China Agricultural Press. (in Chinese)

Wang, L., Guan, Y., Guan, R., Li, Y., Ma, Y., Dong, Z., Liu, X., Zhang, H., Zhang, Y., Liu, Z., & Chang, R. (2006). Establishment of Chinese soybean *Glycine max* core collections with agronomic traits and SSR markers. *Euphytica*, *151*(2), 215–223.

Wien, H. C., & Kueneman, E. A. (1981). Soybean seed deterioration in the tropics. II. Varietal differences and techniques for screening. *Field Crop Research*, *4*(C), 123–132. https://doi.org/10.1016/0378-4290(81)90062-9

Zhang, Q-Y., Li, Y-S., Liu, C-K., Tian, B-W., Tu, B-J., & Mao, J-W. (2015). Key components of eating quality and their dynamic accumulation in vegetable soybean varieties [*Glycine max* (L.) Merr.]. *Acta Agronomica Sinica*, *41*(11), 1692-1700. https://doi.org/10.3724/SP.J.1006.2015.01692

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Nair, R. M., Yan, M., Vemula, A. K., Rathore, A., van Zonneveld, M., & Schafleitner, R. (2023). Development of core collections in soybean on the basis of seed size. *Legume Science*, *5*(1), e158. https://doi.org/10.1002/leg3.158

26396181, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/leg3.158, Wiley Online Library on [08/11/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License