Indian Journal of Agricultural Sciences **92** (2): 164–8, February 2022/Article
https://doi.org/ 10.56093/ijas.v92i2.122129

Creation and validation of core subset of potato (*Solanum tuberosum*) germplasm

DALAMU[1]*, R P KAUR[2], V KUMAR[3], R KUMAR[2], V BHARDWAJ[3] and B CHAUDHARY[4]

ICAR-Central Potato Research Institute, Shimla, Himachal Pradesh 171 001, India

Received: 21 December 2019; Accepted: 02 September 2021

ABSTRACT

Potato (*Solanum tuberosum* L.) is the only non-cereal staple food crop globally. The progress of this temperate crop in the sub-tropical country relies strongly on improvement activities utilizing the genetic resources acquired over a time period. However, proper maintenance, evaluation and utilization of germplasm resources is a cumbersome task. Creation of core set of conserved genetic resources is an efficient and cost effective method of management of conserved germplasm and its use in breeding program. Indian collection of potato (*Solanum tuberosum* ssp *tuberosum*) constitutes 1399 accessions emanating from the five continents maintained under *ex situ* conservation. A core subset of 140 potato germplasm was developed using Power Core software based on 19 morphological traits. Majority of entries in the core subset were from South America (42.14%). Comparison of mean data using Newman-Keul's test and variances using Levene's test for different traits revealed representation of actual diversity of base population by the core subset. Correlations among the traits governed by co-adapted gene complexes were conserved in the core subset. Shannon-Weaver diversity indices indicate that the potato core subset maximized the phenotypic diversity of the potato germplasm. Validation of core subset using principal component analysis reveals that the first 4 PCs represents 55.27% of total variations in core subset compared to 52.50% that in entire collection. The identified core subset can be used in Indian potato improvement programs as well as in genomics studies.

Keywords: Core subset, Potato, Shannon-Weaver diversity index

Potato (*Solanum tuberosum* L.) is a versatile crop cultivated in more than 100 countries located from temperate to subtropics and tropics. It comprises of two subspecies: *tuberosum*; the most cultivated potato known as “Common potato” and *andigena*; majorly cultivated in Central and South America. Potatoes were introduced in India in 17th century by Portuguese sailors or Spaniards. In mid-1800 to mid-1900 systematic introduction of potato germplasm was done. Though a non-native crop, the present potato production is 51.31 million tonnes with average productivity of 24 t/ha. However, this increase is mainly attributed to area expansion rather than yield enhancement (Rana and Anwer 2018). New accessions/germplasms are introduced based on breeder’s requirement. The *tubersoum* germplasm repository constitutes 1399 accessions from 39 donor countries of 5 continents maintained in *ex situ* conservation. However, a very small share of this collection has been utilized in potato breeding programme (Gopal and Oyama 2005). Yields plateau in potato have reached due to confinement of breeders to “working collection”consisting largely of highly adapted material (Kumar *et al.* 2012). Recurrent usage of limited number of adapted genotypes as parental material has led to narrow genetic base of potato varieties worldwide (Plaisted & Hoopes 1989, Wang *et al.* 2019). Historical evidences reveal devastating effects of low levels of genetic diversity eg. Irish famine due to wide cultivation of single potato variety.

Core collection is a subset of entire collection that retains maximum genetic diversity of the species in a region (Brown 1989). Reduction of entire collection to a manageable size under core collection facilitates evaluation for trait of interest especially quantitative traits of economic importance under multi location to identify better parents. Core collections have been developed in potatoes and/or wild potatoes based on morphological, geographical, disease and pest descriptors (Huaman *et al.* 2000), AFLP markers (Bamberg and Alfonso del Rio 2014) and morphological descriptors (Gopal *et al.* 2013). In the present study, a core collection of Indian *tubersoum* germplasm was developed using advanced M strategy of heuristic search (PowerCore)

---

[1] ICAR-Central Potato Research Institute, Kufri-Fagu Unit, Kufri, Shimla, Himachal Pradesh; [2] ICAR-Central Potato Research Institute, RS, Jalandhar, Punjab; [3] ICAR-Central Potato Research Institute, Shimla, Himachal Pradesh; [4] ICAR-Central Potato Research Institute, RS, Modipuram, Meerut, Uttar Pradesh.
*Corresponding author email: dalamu04@gmail.com

based on quantitative descriptors in 1399 accessions maintained at CPRI, Shimla.

MATERIALS AND METHODS

The experimental material comprises of 1399 *Solanum tuberosum* ssp *tubersoum* accessions that were conserved in CPRI field genebank. The experiment was conducted at CPRS, Jalandhar, Punjab research farm (31°02′ N, 75°02′ E, 237 m amsl) in row trial of *rabi* 2015–16. In row length of 3 meter spaced 60 cm apart, each genotype was sown at plant to plant spacing of 20 cm. Standard recommended package and practices of crop management were followed. The nineteen quantitative traits and their method and stage of observations are: Vegetative traits were observed at 75-80 days old crop. Plant height (cm) was recorded as average canopy height of five representative plants from base just above the soil to apical growing tip. Stem per plant was counted based on average number of stems on five representative plants while number of nodes was taken on average number of nodes on main stem of five representative plants. Stem diameter (cm) was recorded as average stem diameter at base just above the soil of five representative plants. The remaining vegetative traits, viz. Petiole length (cm); number of leaflets per compound leaf; number of folioles per compound leaf; leaf length (cm); leaf width (cm); terminal leaflet length (cm); terminal leaflet width (cm); lateral leaflet length (cm); lateral leaflet width (cm) were observed on fully open 4th leaf from the top. Tuber parameters were recorded after dehaulming at 90 days crop duration and subsequent harvesting after 15-20 days. Tuber length (cm) was taken on average tuber length from crown to heel end on five medium-sized representative tubers while tuber width (cm) on the basis of average tuber diameter measured at widest point (centre of tuber). Number of eyes per tuber is average number of eyes on five medium-sized representative tubers. Sprout traits i.e. lightsprout traits were recorded after 30 days of storage of tubers under diffused light at room temperature. The number of sprouts was counted on average number of sprouts measured on five medium-sized representative tubers. Sprout length (cm) and sprout diameter (mm) were measured on average length and average diameter (measured at the base of sprout) of main sprout on five medium-sized representative tubers. The collected data were analyzed using Power Core software (v. 1.0) (http://genebank.rda.go.kr/pwere cpre/) that selects entries or accessions of core sets by the advanced M (maximization) strategy implemented through a modified heuristic algorithm (Kim *et al.* 2007). Mean, range and variances for all quantitative traits in the entire and the core subset were calculated. The means for all traits were compared using Newman-Keul's test (Newman 1939, Keuls 1952). The homogeneity of variances was analyzed using Levene's test. The variable rate (VR %) and the coincidence rate (CR %) were calculated to evaluate properties of the core subset and to test the representativeness of the core subset to the entire collection. Shannon and Weaver (1949) diversity index (H') were used to measure and compare the phenotypic diversity for each trait in the core and entire collection. Phenotypic correlations were calculated among all traits in entire and core subset. Principal component analysis was done to find out relative importance of each trait in capturing the variations in core subset.

RESULTS AND DISCUSSION

A core subset of 140 accessions was developed from 1399 accessions conserved in potato germplasm repository, ICAR-CPRI, Shimla using Power Core method that selects diverse accessions based on high range retention and coefficients of variations. The core subset constitutes 10% of the total accessions conserved. The composition of the core subset suggests predominance of germplasm from South America in the entire collection. Maximum accessions in the core subset (59; 42.14% of core subset) were sampled from South America which is the primary centre of origin of potato and CIP, Peru is also located, which has contributed more than 40% of total potato germplasm imported by India till date. In South America, accessions CP2029,2030, 2037, 2053, 2061, 2085, 2089, 2110, 2165, 2167, 2292, 2305, 2335, 2336, 2368, 2370, 2393, 2399, 2409, 3036, 3142, 3146, 3153, 3181, 3203, 3211, 3222, 3252, 3256, 3258, 3268, 3318, 3337, 3338, 3358, 3362, 3382, 3412, 3413, 3414, 3420, 3421, 3443, 3450, 3491, 3505, 3549, 3577, 3579, 3600, 3608, 3631, 3657, 3668, 3678, 3764, 3766, 3775, 4163 were from Peru while a single accession CP1827 was from Columbia. In remaining accessions, 25% were from North America (35 nos). ie CP1368, 1384, 1390, 1487, 1597, 1730, 1735, 1759, 1762, 1767, 1884, 1907, 1909, 1915, 1918, 1924, 1932, 1940, 3388, 3816, 4095, 4096, 4228 from USA; CP1689, 1698, 1784, 1833, 2354, 3593, 3641, 3649 from Canada and CP1490, 1999, 2015 from Mexico followed by 17.14% from Europe (24 nos), viz. CP1347 from Czechoslovakia; CP1619, 1662 from United Kingdom; CP1449, 1588, 1809, 1826, 1860, 1868,1985, 1988, 4316 from Germany; CP1405, 1418, 1424, 1974, 1978, 3395 from Netherlands; CP2091, 2093 from France; CP3881 from Poland; CP2348, 3880 from Sweden and CP4311 from Denmark, 8.57% from Asia (12 nos) i.e. CP2146, 3424, 3851, 3845, 3853, 3893, 3896, 3901, 3903, 3907 from India while CP3917 and CP2280 from China and Bangladesh, respectively and 0.71% from Australia (1 nos.; CP2285). The accessions from unknown region whose passport information is not available constitute 9 accessions (6.42% of the core subset, viz. CP1143, 1214, 1225, 1235, 1243, 1312, 1335, 1263, 1302). It was observed that the accessions from different locations in the core subset corresponded very well with their actual number in entire germplasm collection.

The range, means and variances of entire and core subset of potato is shown in Table 1. There were non-significant differences between the trait means of the germplasm accessions in the entire and core subset for all the recorded traits (plant height, stems per plant, number of nodes, stem diameter, number of folioles per compound leaf, petiole length, terminal leaflet length, terminal leaflet

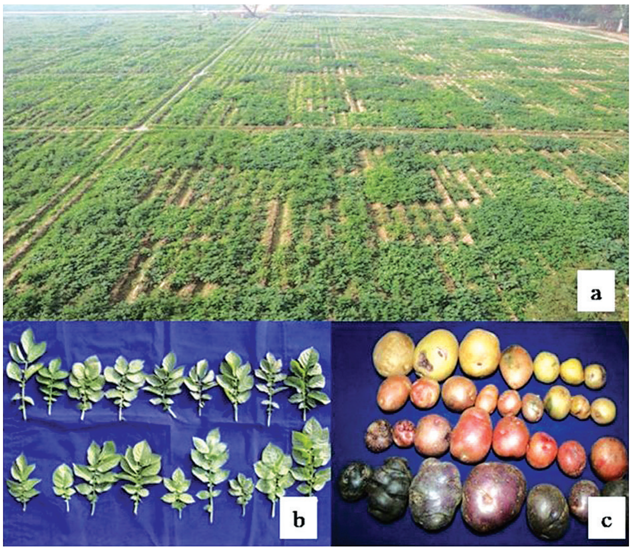

Fig 1 Ariel view of potato germplasm maintenance field site (a); morphological variations depicted in potato leaf (b); and tubers (c).

width, number of leaflets per compound leaf, leaf length, leaf width, lateral leaflet length, lateral leaflet width, tuber length, tuber width, number of eyes per tuber, number of sprouts, sprout length and sprout diameter). Although range was quite high for all the traits in the core subset, but it did not correspond completely with the range observed in the entire collection, except for traits such as number of stems per plant and sprout diameter. Range variation was highest for plant height varying from 12.7 cm to 99.7 cm while least range was recorded for stem diameter ie 6.3-11.9 cm followed by tuber length, lateral leaflet width, lateral leaflet length and tuber width. Variances for traits in the core subset were equal to or higher than that of in the entire collection and were homogenous.

Presence of >80% coincidence rate % and variable rate % is ideal for any created core set as it provides better representation of genetic diversity of base collection. The mean variable rate and coincidence rate were 99.9% and 80.40%, respectively, in the present core subset (Table 2) indicating that the potato core subset was actually representing the phenotypic diversity of entire collection available in Indian Potato genebank. The coincidence rate in the core subset ranged from 57.3 to 100% of the entire collection while the variable rates ranged from 85.3 to 119.8%. Shanon-Weaver diversity indices (H´) measures allelic richness as well as allelic evenness and increase in H´ across traits in core set indicates proper representation

Table 1 Comparisons of range, means and variances of entire and core subset of potato

Trait	Range		Mean*		Variance$			
	Entire	Core	Entire	Core	Entire	Core	F value	Probability
PH	12.7-120	12.7-99.7	58.0a	58.1a	201.92	240.57	1.66	0.198
SPP	1-10	1-10	3.6a	3.5a	1.06	1.04	<0.01	0.960
NN	3.3-22	3.3-18.7	10.6a	10.7a	6.16	7.39	1.77	0.184
StD	4.7-12.3	6.3-11.9	8.4a	8.5a	1.02	1.01	0.35	0.553
NFPCL	2-33	4-29.3	10.9a	10.2a	15.84	19.91	1.90	0.168
PL	0.8-9.3	0.8-5.7	2.8a	2.8a	0.39	0.51	2.75	0.097
TLlL	2.7-11.2	3.7-11.1	6.3a	6.3a	1.35	1.39	0.18	0.668
TLlW	1.7-13.8	2.2-11.6	3.8a	3.8a	1.07	1.05	<0.01	0.999
NLlPCL	4.3-15.7	4.7-15.7	7.8a	7.6a	3.12	3.5	0.48	0.490
LL	5.6-25.4	7.4-23.2	15.7a	15.5a	8.37	8.09	0.02	0.896
LW	3.5-17.1	3.5-14.4	9.0a	8.9a	3.71	2.79	3.53	0.061
LLlL	2.4-8.5	3.1-8.4	5.2a	5.1a	0.83	0.68	0.34	0.561
LLlW	1.2-5.7	1.8-4.6	3.0a	3.0a	0.41	0.28	0.06	0.810
TL	3.1-11.1	4.1-9.9	6.4a	6.4a	0.7	0.66	0.11	0.744
TW	2.5-7.8	2.5-7.4	5.0a	5.0a	0.29	0.3	1.49	0.223
NEPT	2.2-17.6	2.4-12	7.6a	7.6a	2.26	2.03	0.03	0.859
NS	2.7-16	3.3-14	7.7a	7.8a	2.39	2.07	0.34	0.562
SL	0.1-3.7	0.1-2.8	0.9a	0.8a	0.19	0.16	0.44	0.505
SD	0.4-12.7	0.4-12.7	5.8a	5.7a	4.12	3.43	0.94	0.334

PH: Plant Height; SPP: Stem Per Plant; NN: Number of nodes; StD: Stem Diameter; NFPCL: Number of Folioles Per Compound Leaf; PL: Petiole Length; TLlL:Terminal Leaflet Length; TLlW:Terminal Leaflet Width; NLlPCL: Number of Leaflets Per Compound Leaf; LL: Leaf Length; LW: Leaf Width; LLlL: Lateral leaflet Length; LLlW: Lateral leaflet Width; TL: Tuber Length; TW: Tuber Width; NEPT: Number of Eyes Per Tuber; NS: Number of Sprouts; SL: Sprout Length; SD: Sprout Diameter.

[*], means of entire and core were tested using Newman-Keul's test, means followed by similar letter were non-significant at P=0.05;

[$], variances of entire and core were tested using Levene's test and were non-significant for all traits at P=0.05.

Table 2 Variable rate (VR), coincidence rate (CR) and Shanon-Weaver diversity indices (H index) of entire and core subset of potato

Trait	VR (%)	CR (%)	H index			
			Entire	Core	Mean	SE±
PH	108.9	81.1	0.627	0.609	0.618	0.009
SPP	102.3	100.0	0.631	0.580	0.606	0.025
NN	108.0	82.1	0.624	0.606	0.615	0.009
StD	98.6	73.6	0.623	0.615	0.619	0.004
NFPCL	119.8	81.7	0.615	0.550	0.582	0.033
PL	115.8	57.3	0.613	0.581	0.597	0.016
TLlL	100.5	86.3	0.636	0.596	0.616	0.020
TLlW	98.2	77.7	0.517	0.493	0.505	0.012
NLlPCL	107.6	97.1	0.532	0.560	0.546	0.014
LL	100.0	79.8	0.630	0.610	0.620	0.010
LW	88.2	80.4	0.614	0.597	0.605	0.009
LLlL	91.2	87.3	0.627	0.621	0.624	0.003
LLlW	85.3	62.7	0.616	0.614	0.615	0.001
TL	96.5	72.3	0.633	0.629	0.631	0.002
TW	101.4	93.0	0.612	0.580	0.596	0.016
NEPT	94.4	62.3	0.609	0.627	0.618	0.009
NS	92.2	80.0	0.629	0.609	0.619	0.010
SL	96.9	73.4	0.574	0.572	0.573	0.001
SD	92.6	100.0	0.597	0.582	0.589	0.007
Mean	99.9	80.4	0.608	0.591	0.600	
SE±	2.07	2.77	0.008	0.007	0.007	

of available diversity in the entire collection. A low H´ indicates unbalanced frequency of classes for a trait and a lack of genetic diversity. Mean Shanon-Weaver diversity indices was same for both entire (0.600 ±.008) and core subset (0.59±0.007) representing that the allelic diversity of the entire collection represented by phenotypic diversity was included or retained in the core subset. In Andigena potato core collection Gopal *et al.* (2013) also reported edge of heuristic search over other methods for development of core set. Terminal leaflet width has lowest H' value while tuber length has highest value in core subset. Lateral leaflet width and sprout length have similar H' estimates in both core and entire collection.

Phenotypic correlations were estimated between all 19 quantitative traits in the entire and core collections, respectively. Correlations pattern was same in the entire and core representing preservation of associations of entire collection in the core subset and signifies role of co-adapted gene complexes governing these traits. However, in comparison to 115 significant correlation coefficients in entire collection, only 69 were significant (P=0.05) in core subset which may be due to sampling effects (Gangopadhyay *et al.* 2010). The correlation coefficients in core subset range from 0.167 to 0.839 (Supplementary Table 1). However, correlation coefficients with an absolute value >0.71 or ≤ -0.71 are significant (Skinner *et al.* 1999) as it capture ≥50% variation in one trait as predicted by the other. Definite correlations among traits like lateral leaflet length, lateral leaflet width, leaf width, leaf length, sprout length, sprout diameter, plant height, number of nodes, tuber length, tuber width, number of folioles, number of leaflets per compound leaf, terminal leaflet length and terminal leaflet width, that were recorded in entire collection were conserved in core subset.

Principal component analysis (PCA) evaluates interrelationships among the different traits (Data not presented). The first four principal components represent 55.27% of total variation in comparison to 52.50% in entire collection defining the extent of variations represented by the core. The cumulative variance of 55.27% by the first four axes with Eigen value of >1.0 indicates that the identified traits within the axes exhibited great influence on the phenotype of accessions. The first principal component (PC1) explains 24.34% of the total variation and was attributed by traits, viz. terminal leaflet length, terminal leaflet width, leaf length, leaf width, petiole length, lateral leaflet length, tuber width and lateral leaflet width. PC2 explained 12.39% of the total variation and was attributed to plant height and number of nodes while PC3 and PC4 explains 9.54 and 9.00% of total variations and traits sprout length, sprout diameter and number of eyes per tuber differentiates the genotypes.

Extensive evaluation of entire collection for trait of interest requires huge investment and is quite cumbersome. Development of core set provides an easy access point to genetic resources without losing the actual diversity. Core collections are most suitable genetic resources for discovering new sources of variations to use in crop improvement activities. Core set drastically reduces the number of accessions to be evaluated and immensely facilitates extensive evaluations across locations. The accessions forming a part of core set would be potential parental lines for breeding programs, holding a much larger representative diversity of genes. Further, molecular characterization of core set may be undertaken to define the extent of variations in the base collection and reaffirm the morphological findings. The identified potato core subset in the present study may be revised periodically as and when new accessions are added to the collection and characterization information become available.

REFERENCES

Bamberg J and del Rio A.2014. Selection and validation of an AFLP marker core collection for the wild potato *Solanum microdontum*. *American Journal of Potato Research* **91**: 368–75.

Brown AHD 1989. Core collections: a practical approach to genetic resources management. *Genome* **31**(2): 818–24.

Gangopadhyay K K, Mahajan R K, Kumar G, Yadav S K, Meena B L, Pandey C and Bisht I S. 2010. Development of a core set in brinjal. *Crop Science* **50**: 755–62.

Gopal J and Oyama K. 2005.Genetic base of Indian potato selections as revealed by pedigree analysis. *Euphytica* **142**:

23–31.

Gopal J, Kumar V, Kumar R and Mathur P. 2013. Comparison of different approaches to establish a core collection of andigena (*Solanum tuberosum* group *andigena*) potatoes. *Potato Research* **56**: 85–98.

Huaman Z, Ortiz R and Gomez R. 2000. Selecting a*S. tuberosum sub sp. andigena* core collection using morphological, geographical, disease and pest descriptors.*American Journal of Potato Research* **77**: 183–90.

Keuls M. 1952. The use of the "Studentized range" in connection with an analysis of variance. *Euphytica* **1**: 112–22.

Kim K W, Chung H K, Cho G T, Ma K H,Chandrabalan D, Gwag J G, Kim T S, Cho E G and Park Y J. 2007. PowerCore: a program applying the advanced M strategy witha heuristic search for establishing core sets. *Bioinformatics* **23**: 2155–62.

Kumar R, Gopal J and Pandey S K. 2012.Genetic improvement for yield and tuber size in Andigena potatoes (*Solanum tuberosum* sub sp *andigena*) after one cycle of recurrent selection.*Indian Journal of Agricultural Sciences* **82**(10): 885–8.

Newman D. 1939. The distribution of range in samples from a normal population expressed in terms of an independent estimate of standard deviation. Biometrika **31**: 20–30.

Plaisted R L and Hoopes R W. 1989. The past record and future prospects for the use of exotic potato germplasm. *American Potato Journal* **66**: 603–27.

Rana R K and Anwer M D E.2018. Potato production scenario and analysis of its total factor productivity in India. *Indian Journal of Agricultural Sciences* **88** (9): 1354–61.

Shannon C E and Weaver W. 1949. The mathematical theory of comunication. University of Illinois Press, Urbana.

Skinner D Z, Bauchan G R, Auricht G and Hughes S.1999. A method for the efficient management and utilization of large germplasm collections. *Crop Science* **39**: 1237–42.

Wang Y, Rashid MAR, Li X, Yao C, Lu L, Bai J, Li Y, Xu N, Yang Q, Zhang L, Bryan GJ, Sui Q and Pan Z. 2019. Collection and evaluation of genetic diversity and population structure of potato landraces and varieties in China. *Frontiers in Plant Sciences* **10**: 139.