

# Understanding genetic diversity, population structure and development of a core collection of Indian accessions of watermelon (*Citrullus lanatus* (Thunb.) Matsum. and Nakai)

Saheb Pal<sup>1</sup>, Muttanna Revadi<sup>1</sup>, RN Thontadarya<sup>1</sup>, DC Lakshmana Reddy<sup>2</sup>, B. Varalakshmi<sup>1</sup>, Chithra Devi Pandey<sup>3</sup> and Eguru Sreenivasa Rao<sup>1\*</sup> 

<sup>1</sup>Division of Vegetable Crops, <sup>2</sup>Division of Biotechnology, ICAR-Indian Institute of Horticultural Research, Bangalore, India and <sup>3</sup>Division of Germplasm Conservation, ICAR-National Bureau of Plant Genetic Resources, New Delhi, India

Received 29 May 2020; Accepted 6 October 2020 – First published online 4 November 2020

## Abstract

Most of the modern-day improved watermelon varieties succumb to various biotic and abiotic stresses mainly because of their narrow genetic base. Insights into the genetic diversity and population structure are crucial for broadening the genetic base and improving the adaptive value. The present experiment was conducted to study the genetic diversity and population structure of a germplasm panel comprising 336 *Citrullus* sp. accessions. Another objective was to formulate a core collection of Indian *Citrullus* sp. accessions. Data from 23 highly polymorphic microsatellite markers were used for genetic diversity and population structure analysis while both molecular and phenotypic data from 17 traits were used to formulate the core set. The markers yielded a total of 69 alleles with an average of three alleles per locus. Initially, the accessions clustered into two populations and an admixture group. Intra-population analysis revealed three and two statistically distinct subpopulations in Pop I and Pop II, respectively. The exotic collections were predominant in Pop I-A, Pop II-A and Pop II-B while the Indian accessions were preponderant in Pop I-B and Pop I-C. Pop I-B recorded the maximum magnitude of gene diversity and the highest number of alleles. The well-adapted Indian landraces could be deployed in future watermelon improvement programmes. The formulated core collection ( $n = 46$ ; 23.71% of the entire collection studied) would ease in maintenance of the diversity present among indigenous *Citrullus* sp. accessions; would ease trait search while exploring Indian diversity and can be pooled with other collection(s) to form a global core of watermelon.

**Keywords:** Genetic diversity, Indian core set, Population structure, Watermelon

## Introduction

Watermelon (*Citrullus lanatus* (Thunb.) Matsum. & Nakai) is a speciality vegetable valued mainly for its fruits and seeds. The genus *Citrullus* consists of cultivated watermelon and six other distinct species *viz.*, *C. amarus*,

*C. mucospermus*, *C. colocynthis*, *C. ecirrhosus*, *C. rehmii* and *C. naudinianus* (Chomicki and Renner, 2015). The crop is native to Africa (Renner *et al.*, 2017). However, considerable diversity exists in India and the species *C. colocynthis* is found to grow in the wild in the country (Dane *et al.*, 2007). Over 4000 years ago, domestication of *C. lanatus* started in North-eastern Africa, Egypt and Sudan from which, few elite selections were introduced into India during the 8<sup>th</sup> Century (Paris *et al.*, 2013). Since

\*Corresponding author. E-mail: [esrao1973@gmail.com](mailto:esrao1973@gmail.com)

then, it is under cultivation in irrigated as well as conserved moisture conditions and mainly for seed purpose in the arid and semi-arid regions of Rajasthan. As a result of gene flow between *C. lanatus* and *C. colocynthis*, several well-adapted landraces with novel characters have evolved in India (Mahla *et al.*, 2014). During the 11<sup>th</sup> Century, the crop was introduced into China (Zhao, 2015), where few local dessert varieties and an edible-seeded variety were developed and are still prevalent (Levi *et al.*, 2017). Owing to the presence of ample genetic diversity, the area adjoining India and China is considered as the secondary centre of origin of the genus *Citrullus* (Wehner, 2008). The crop was introduced into the Americas after 1493 (Zoltán *et al.*, 2007) and several important cultivars have been developed over the last 250 years (Levi *et al.*, 2017). Some of these US cultivars like Mickey Lee and Bush Sugar Baby share genetic relatedness to the Asian gene pool (Hwang *et al.*, 2011).

There exists ample variability for different horticultural traits, but several studies indicated narrow genetic diversity among the cultivated varieties (Kwon *et al.*, 2010; Minsart *et al.*, 2011; Levi *et al.*, 2013). The narrow genetic base of the cultivated species might be the result of 'founder effect' originating from the introduction of few superior genotypes and their use in genetic improvement and directional selection for few consumer appealing traits (Dane and Liu, 2007). Further, the narrow genetic base poses a challenge for breeders, especially for resistance breeding. A large germplasm panel without the information of trait diversity hinders their successful utilization (Zhang *et al.*, 2016). This necessitates a thorough understanding of genetic diversity in the available germplasm for broadening the genetic base using wild accessions.

A core set (Frankel and Brown, 1984) represents maximum diversity with minimum repetitiveness, thus helps in efficient management, conservation and utilization of germplasm. To date, a core watermelon collection has been established by Zhang *et al.* (2016) from a germplasm panel including only 51 Indian accessions, which might not represent the extent of diversity present in Indian germplasm. Keeping the above facts in mind, we attempted to study the genetic diversity and population structure of Indian *Citrullus* sp. accessions in comparison with exotic collections and to formulate a core set of Indian accessions of watermelon.

## Materials and methods

### Plant materials

The experimental material comprised of 336 genotypes (online Supplementary Table S1) collected from 15 different countries (194 inbreds and 19 hybrids and their derivatives from India, 77 from the USA, 15 from Nigeria, eight

from Thailand, five from Senegal, three each from China and Zimbabwe, two each from Japan, South Africa, Mali and Botswana and one each from Ghana, Moldova, Taiwan and Zambia). The germplasm panel consisted genotypes of the cultivated species (*C. lanatus*;  $n = 241$ ) along with three other related-wild species *viz.*, *C. mucospermus* ( $n = 49$ ), *C. colocynthis* ( $n = 44$ ) and *C. amarus* ( $n = 2$ ) were used for the current study (online Supplementary Table S2). The genotypes were collected from the respective countries through germplasm explorations while the genotypes mentioned under the USA were collected from Germplasm Resource Information Network by ICAR-NBPGR (online Supplementary Table S3). All these genotypes are actively being maintained at ICAR-IIHR.

### Experimental site and morphological evaluation

The genotypes were evaluated phenotypically during the *rabi* season of 2015–2016 in Randomized Complete Block Design with three replications comprising 10 plants in each replication. The site is located at 890 m above mean sea level at 13°08'26.6"N latitude and 77°30'2.2"E longitude. The soil in the experimental plot has loamy-sand texture, neutral pH (6.7) and belongs to Alfisol order. The total nitrogen, phosphorus and potassium content in the soil are 188.2, 89.6 and 350 kg/ha; estimated following the standard alkaline permanganate (Subbiah and Asija, 1956); sodium bicarbonate (Olsen *et al.*, 1954) and flame-photometric (Black, 1965) methods, respectively. Fourteen-day old seedlings were transplanted in the experimental plot at a spacing of 180 cm × 45 cm. Standard cultural practices were followed to raise a healthy crop stand except for fungicidal sprays (including tebuconazole, pyraclostrobin and trifloxystrobin) which are generally used for managing gummy stem blight in watermelon. Observations were recorded for 12 qualitative characters *viz.*, early plant vigour, plant growth habit, degree of lobing of the leaf blade, sex expression, fruit shape, primary and secondary rind colour, rind stripe colour, flesh colour, distribution of grooves on the fruit skin, flesh bitterness, seed-coat colour, stripe colour and five quantitative traits *viz.*, fruit weight, vine length, number of primary branches per plant, number of fruits per plant and Total Soluble Solids as per the minimal descriptor for watermelon of ICAR-National Bureau of Plant Genetic Resources, New Delhi, India (online Supplementary Table S3). For the vegetative and fruit characters, observations were recorded on seven random plants per replication for each accession.

### Simple sequence repeat genotyping for molecular analysis

Three weeks after transplanting, bulked leaf samples were collected from 10 random plants from each accession for

molecular analysis. Genomic DNA was isolated from these samples by following a modified CTAB method (Doyle and Doyle, 1987) with minor modifications. Air-dried DNA pellets were re-suspended in 1X TE Buffer (pH=8.0) and were quantified by using UV-Spectrophotometer (GeneQuantPro, GE Healthcare, USA). The dilution was made accordingly in milli-Q water to obtain the final working concentration of 10 ng DNA/ $\mu$ l.

Zhang *et al.* (2012) identified 23 simple sequence repeat (SSR) markers which possessed high polymorphism information content (PIC) and could consistently group watermelon accessions similar to the grouping pattern made using SNP data obtained from whole-genome sequencing as well as using data on morphological traits. Thus, these markers were considered to be a core set for diversity analysis. Among them, 21 markers were included in the current study. The remaining two markers *viz.*, BVWS00734 and BVWS00333 were yielding faint inconclusive bands. Therefore, we replaced these markers with BVWS01708 and BVWS00539 which fall in the same bins respectively, on the linkage map of Ren *et al.* (2012). The SSR primers were synthesized at Bioserve India Pvt. Ltd., Hyderabad. All the 336 accessions were genotyped using these 23 markers in a thermal cycler (Eppendorf, USA) with the following conditions: initial denaturation at 94°C for 5 min; denaturation 94°C for 30 s, annealing at 55°C for 30 s, extension at 72°C for 90 s for 35 cycles followed by a final extension at 72°C for 8 min. Each reaction (volume 16  $\mu$ l) consisted of 4  $\mu$ l nuclease-free water, 3  $\mu$ l of genomic DNA (diluted to 10 ng/ $\mu$ l), 2  $\mu$ l of each of dNTPs (1 mM), forward and reverse primers (10 pM), 2.5  $\mu$ l of 10X polymerase chain reaction (PCR) buffer with MgCl<sub>2</sub> and 0.5  $\mu$ l *Taq* DNA polymerase (Genei, Bangalore, India). The PCR amplicons were run in 4% molecular biology grade agarose gel stained with 0.5  $\mu$ g/ml of ethidium bromide (Genei, Bangalore, India) and was visualized in the gel documentation system (UVI-TEC, Cambridge). For each marker, the amplicons with the smallest size were scored as 1 and subsequently, incremental values were assigned with the increase in size. The missing values were scored as -9.

### Statistical analysis

Major allele frequency, the average number of alleles, observed and expected heterozygosity ( $H_o$  and  $H_e$ ) and PIC for each SSR marker were computed using PowerMarker v. 3.25 (Liu and Muse, 2005).

The genotypic data of the SSR markers were used for population structure analysis employing admixture model-based approach implemented in STRUCTURE software v.2.3.4 (Pritchard *et al.*, 2000). The data were tested from  $K = 1-10$  populations ( $K =$  putative number of populations) with three replications and burn-in of 30,000 followed by

1,00,000 MCMC repeats for each run. Later, the three replicates of the analysis were integrated using CLUMPP software (Jakobsson and Rosenberg, 2007). To avoid overestimation of the number of populations estimated from log-probability [LnP(D)] data, the number of subpopulations was estimated based on  $\Delta K$ , by using ad-hoc web-based programme STRUCTURE HARVESTER (<http://taylor0.biology.ucla.edu/structureHarvester/>) (Evanno *et al.*, 2005). A membership probability of  $\geq 0.75$  was considered for assigning a genotype into a population and the genotypes with membership probability less than this value were assigned to an admixture group as per Rao *et al.* (2012). The genotypic data of individuals in each population were further subjected to intra-population analysis following the same protocol (Pritchard *et al.*, 2000). Unweighted neighbour-joining tree for the *Citrullus* sp. accessions was also constructed with the genotypic data using DARwin v. 6.0.21 (Perrier and Jacquemoud-Collet, 2006).

The core collection was developed employing PowerCore v. 1.0 software which is based on a modified heuristic algorithm (Kim *et al.*, 2007). For this purpose, the data from 17 phenotypic traits and 23 microsatellite markers were utilized. To assess the homogeneity of the core with the original collection, Shannon Weaver diversity index (Shannon and Weaver, 1963) and Nei diversity index (Nei, 1973) were estimated for both the phenotypic traits and the SSR markers in both collections. Coverage of the phenotypic traits of the original collection was also estimated in the core as suggested by Kim *et al.* (2007).

## Results

### Phenotypic and molecular evaluation of the panel

Wide genetic variation was observed across the genotypes for 12 qualitative and five quantitative traits (online Supplementary Table S1). All the genotypes, except IIHR-86 and IIHR-184, had moderate to good early plant vigour. Two Indian inbreds *viz.*, IIHR-15 and IIHR-86 had bushy plant habit. Further, 72 accessions had andromonoecious sex form and 17 accessions had nonlobed leaves. The maximum number of primary branches per plant was noted in IIHR-515 (*C. mucospermus*), while it was minimum in IIHR-277 (*C. lanatus*). IIHR-259 was most vigorous in terms of length of the vines, while IIHR-86 had the shortest vine. The fruit shape varied from round to oblong while the primary fruit skin colour varied from light green to yellow, the former states being predominant for both the traits. Most of the genotypes possessed medium to dark green striped fruits with few exceptions. The flesh colour in different *C. lanatus* genotypes varied as red, pink, canary yellow and salmon yellow. However, all the genotypes belonging to *C. colocynthis*, *C. amarus* and

**Table 1.** General statistics of the SSR markers computed using PowerMarker v. 3.25

Marker	Major Allele frequency	Number of alleles	Gene diversity	Heterozygosity	PIC
BVWS00948	0.5092	2	0.4998	0.8646	0.3749
BVWS00155	0.8763	3	0.2208	0.0108	0.2036
BVWS00297	0.5912	2	0.4834	0.0943	0.3665
BVWS00314	0.7759	2	0.3478	0.0658	0.2873
BVWS00048	0.6689	3	0.4926	0.0131	0.4377
BVWS00208	0.5901	4	0.5252	0.2453	0.4364
BVWS01708	0.5655	3	0.4929	0.2655	0.3731
BVWS00441	0.7417	3	0.3848	0.1457	0.3133
BVWS00658	0.5209	2	0.4991	0.0941	0.3746
BVWS00106	0.5926	4	0.5886	0.1235	0.5444
BVWS00233	0.6315	3	0.4688	0.0031	0.3631
BVWS00897	0.6214	3	0.5403	0.1197	0.4811
BVWS00433	0.4643	3	0.6408	0.0750	0.5686
BVWS00170	0.7169	3	0.4402	0.1490	0.3925
BVWS00369	0.6810	4	0.4880	0.2055	0.4427
BVWS00826	0.5922	2	0.4830	0.0863	0.3664
BVWS00209	0.5227	4	0.6252	0.2447	0.5658
BVWS00539	0.9207	2	0.1460	0.0032	0.1353
BVWS00843	0.5625	2	0.4922	0.1453	0.3711
BVWS02048	0.5232	3	0.5146	0.1207	0.3986
BVWS00236	0.5505	3	0.5030	0.0917	0.3854
BVWS00228	0.5128	5	0.5691	0.1090	0.4801
BVWS00839	0.5681	4	0.5170	0.0036	0.4134
Mean	0.6217	3	0.4767	0.1426	0.3946

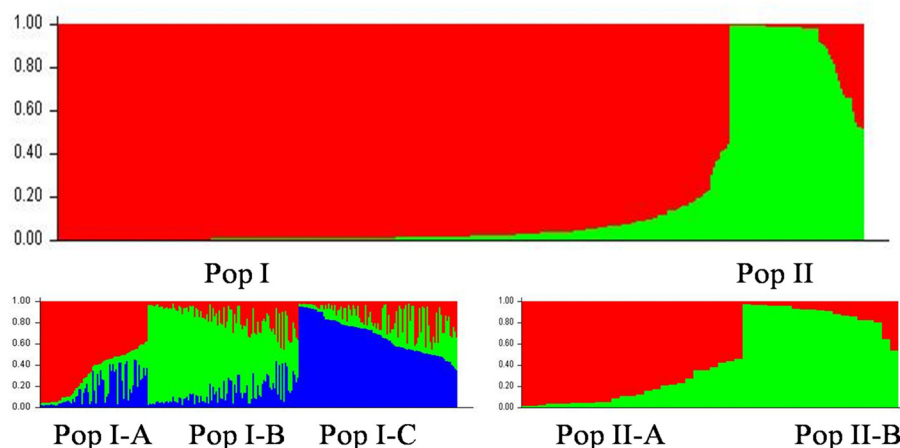
*C. mucospermus* had white-fleshed fruits. The average fruit weight varied from 0.1 kg in IIHR-573 (*C. colocynthis*) to 4.54 kg in IIHR-609 (*C. lanatus*) while the average number of fruits varied from 0.20 in 23 different accessions to 3.75 in IIHR-197 (*C. colocynthis*). Most of the genotypes of *C. lanatus* and *C. colocynthis* were dark-brown to tan seeded while all the *C. mucospermus* accessions were white seeded. The total soluble solid content varied from 1.0°Brix in IIHR-582 (*C. colocynthis*) to 11.33°Brix in IIHR-608. In addition to the *C. lanatus*, six *C. colocynthis* viz., IIHR-158, IIHR-202, IIHR-272, IIHR-286, IIHR-299 and IIHR-568 and 11 *C. mucospermus* viz., IIHR-310, IIHR-517, IIHR-530, IIHR-532, IIHR-533, IIHR-533, IIHR-534, IIHR-536, IIHR-537, IIHR-556, IIHR-587 and IIHR-605 accessions produced non-bitter fruits.

The 23 SSR markers yielded a total of 69 alleles across the germplasm panel. The range for the number of alleles varied from 2 (BVWS00948, BVWS00297, BVWS00314, BVWS00658, BVWS00826, BVWS00539 and BVWS00843) to 5 (BVWS00228) with an average of 3 alleles per locus. Average gene diversity and PIC were 0.4767 and 0.3946

respectively, recording mean heterozygosity of 14.26% over the whole panel under study (Table 1).

### Population structure analysis

The model-based analysis implemented in STRUCTURE resulted in 30 data sets (three replications for each  $K$  from 1 to 10), which were then permuted for each  $K$  value by using CLUMPP software. The actual number of populations could not be inferred from the [LnP(D)] value as it increased continuously for each  $K$  from 1 to 10 (online Supplementary Figure S1A). Hence, second-order statistics (Evanno *et al.*, 2005) was applied and a sharp peak of  $\Delta K$  was obtained at  $K = 2$ , suggesting the presence of two major populations (online Supplementary Figure S1B). The Population I (hereafter referred as Pop I) consisted of 271 accessions, majority of which were from India (72.69%,  $n = 197/271$ ), followed by genotypes from the USA (55), Thailand (7), China (3), Nigeria (3), Japan (2) and one genotype each from Taiwan, Senegal, Botswana and Moldova. The



**Fig. 1.** Model-based clustering of 336 accessions in two main populations and five subpopulations.

population II (hereafter referred as Pop II) contained 46 accessions from the USA (18), Nigeria (10), India (7), Zimbabwe (3), two each from South Africa and Mali, and one each from Thailand, Senegal, Ghana and Botswana (Fig. 1, online Supplementary Table S3). The remaining 19 genotypes (nine from India, four from the USA, three from Senegal, two from Nigeria and one from Zambia) had membership probabilities less than 0.75 for any of the populations, hence were classified into an admixture group. The unweighted neighbour-joining tree was also constructed for the accessions (Fig. 2) and the clades were highlighted based on their area of the collection (A) and species (B).

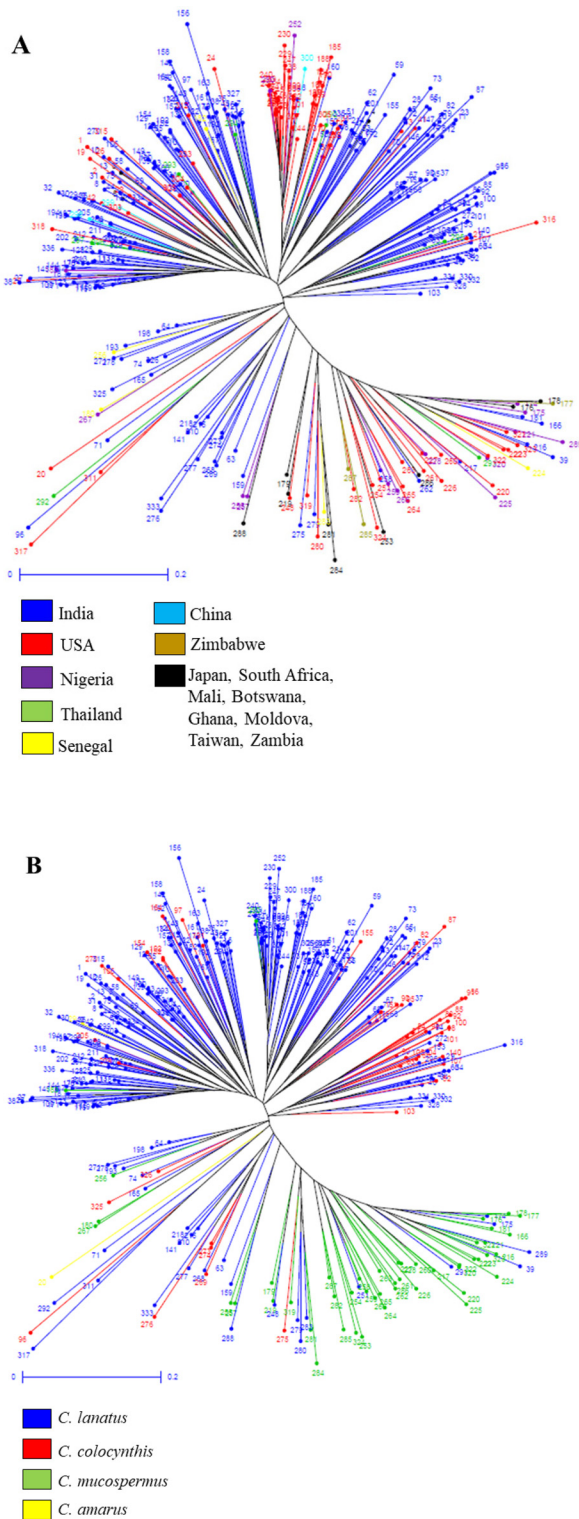
The two main populations could further be subdivided into five subgroups based on the  $\Delta K$  values 3 and 2, for Pop I and Pop II, respectively. Population I, predominated by *C. lanatus* accessions (83.76%), consisted of three subgroups *viz.*, Pop I-A with 31 accessions (26 from the USA, three from India and one each from China and Nigeria); Pop I-B with 55 accessions (50 from India, three from USA and one each from Thailand and Botswana) and Pop I-C with 54 accessions (46 from India, five from USA and one each from Nigeria, Thailand and Taiwan) and an admixture group with 131 genotypes, predominantly from India (74.81%,  $n = 98/131$ ) (online Supplementary Table S3). Population II was subdivided into two subgroups *viz.*, Pop II-A and Pop II-B, consisting of 21 and 17 genotypes respectively, and an admixture group with eight accessions. Both these sub-populations were predominated by the *C. mucospermus* accessions (90.47% and 94.12%; respectively; Fig. 1 and online Supplementary Table S2).

Analysis of Molecular Variance depicted significant differences among the five sub-populations, with a maximum genetic diversity of 33.5%. The genetic diversity within Pop I-B was 22.04% followed by Pop I-C (14.75%), Pop I-A

(5.94%), Pop II-A (5.65%) and Pop II-B (4.25%). Pop I-B had maximum genetic diversity in terms of maximum values of both PIC value (0.6396) and the number of alleles per locus (2.5652) with a moderate frequency of major alleles (0.6704) while the minimum PIC (0.1843) and the maximum frequency of major allele (0.8455) was recorded in Pop I-A. Pop II-B exhibited the minimum number of alleles per locus (2.0435). Both expected and observed heterozygosity were maximum in Pop II-A (0.3286 and 0.1689, respectively) whereas the values were minimum in Pop I-A (0.2167 and 0.0981 respectively; Table 2). The genetic distance between these subpopulations, revealed by the pairwise  $F_{st}$  values (Table 3), indicated Pop I-A and Pop II-B to be the most diverse ( $F_{st} = 0.6021$ ) followed by Pop I-C and Pop II-B ( $F_{st} = 0.5558$ ). The genetic distance ( $F_{st} = 0.2312$ ) was found to be the minimum between Pop I-B and Pop II-A (Tables 2 and 3).

### Formulation of the core set

A total of 194 indigenous *Citrullus* inbreds (19 hybrids and their derivatives) were excluded from a total of 213; hereafter referred as the entire collection), belonging to five different species were subjected to core set analysis. The heuristic search in PowerCore identified 46 accessions as the core collection (online Supplementary Table S1). The values of MD, VD, CR and VR for the core set were 3.35, 29.89, 95.13 and 116.84%, respectively, as compared with the entire collection of Indian inbreds. As the VR was more than 100%, the resultant core collection is a good representative of the entire collection. The CR was also more than the 80% threshold suggested by Kim *et al.* (2007). This indicated a homogeneous distribution of the genotypes in the core collection. The distribution of Nei gene indices and Shannon Weaver diversity indices were similar in core and



**Fig. 2.** Genetic relationship among 336 *Citrullus* sp. accessions. Unweighted neighbour-joining tree constructed using the genotypic data obtained from 23 SSR markers across the accessions. Clades are highlighted based on their area of collection (A) and species (B).

the original collection (Fig. 3). The value for Nei diversity index varied from 0.078 (growth habit) to 0.779 (BVWS00106) with an average of 0.547 in the entire population, while it varied from 0.126 (growth habit) to 0.808 (flesh colour) with an average of 0.587 in the core collection. Similarly, in the base collection, Shannon Weaver diversity index varied from 0.078 (growth habit) to 1.752 (BVWS00106) with an average of 1.016 in the entire collection, whereas in the core collection, it varied from 0.146 (growth habit) to 1.755 (BVWS00106) with an average of 1.120. A total of 16 markers (69.56% of 23 SSRs) had Nei diversity index of more than 60%, which implied the presence of greater genetic diversity in the accessions across the markers (Fig. 3).

## Discussion

### *Phenotypic and molecular evaluation of the panel*

Watermelon is a speciality vegetable crop which is nowadays bred keeping in view the specific traits preferred by the growers, traders and consumers. In general, a variety with a greater number of fruits per plant and bigger size of the fruits are widely accepted by the growers; crispy sweet flesh with fewer seeds by consumers and good shipping of varieties by traders. Bushiness of the vine is the most desirable target trait when watermelon is bred for container or rooftop garden or for planting in high density. The non-lobed leaf trait can be used as a morphological marker for maintenance of parental lines of hybrids as the character is recessive to normal lobed leaf. This is governed by a single dominant gene (*CILL1*) situated on chromosome 4 (Wei *et al.*, 2017). Though sweetness and flesh colour are the two most important traits that strongly influence consumer preference (Kyriacou *et al.*, 2017); fruit shape, stripe pattern and as well as seed colour also stands as one of the key acceptance determining traits.

Microsatellite loci are prone to replication errors leading to multiple alleles. Therefore, their use in the characterization of genetic diversity provides better insights as compared to SNPs which are generally biallelic (Reddy *et al.*, 2015). The average PIC value (0.3946) of all the SSR markers across the studied germplasm was less in magnitude than that reported by Mujaju *et al.* (2011) (PIC=0.92 for 22 Southern African accessions), Kwon *et al.* (2010) (PIC=0.50 for 49 commercial watermelon cultivars) and Zhang *et al.* (2016) (PIC=0.49 for 1197 accessions belonging to seven *Citrullus* species), Singh *et al.* (2017) (PIC=0.46 for 88 accessions belonging to three *Citrullus* species). However, our results corroborate with the PIC value observed by Nantoume *et al.* (2013) (PIC=0.35 for 134 landraces from Mali) and Munisse *et al.* (2013) (PIC=0.34 for 96 genotypes from Mozambique). Among

**Table 2.** Genetic diversity in the five sub-populations based on 23 SSR markers

Population	Major allele frequency	Genotype number	Sample size	Average Allele number	Gene diversity	Heterozygosity	PIC
Overall	0.6217	5.2174	336	3.0000	0.4767	0.1426	0.3946
Pop I overall	0.6551	4.5652	271	2.8696	0.4402	0.1362	0.3623
Pop I-A	0.8455	2.4348	31	2.0870	0.2167	0.0981	0.1843
Pop I-B	0.6704	3.913	55	2.5652	0.4184	0.1533	0.6396
Pop I-C	0.7828	3.1739	54	2.3913	0.3046	0.1194	0.2580
Pop II overall	0.7225	3.5217	46	2.4783	0.3553	0.1631	0.3010
Pop II-A	0.7551	2.9565	21	2.3478	0.3286	0.1689	0.2789
Pop II-B	0.7942	2.5652	17	2.0435	0.2750	0.1222	0.2324

**Table 3.** Pair-wise  $F_{st}$  estimates among the five subpopulations identified

Population	Pop I-A	Pop I-B	Pop I-C	Pop II-A	Pop II-B
Pop I-A	0.0000	0.3661	0.4443	0.4716	0.6021
Pop I-B		0.0000	0.2534	0.2312	0.3877
Pop I-C			0.0000	0.4333	0.5558
Pop II-A				0.0000	0.2383
Pop II-B					0.0000

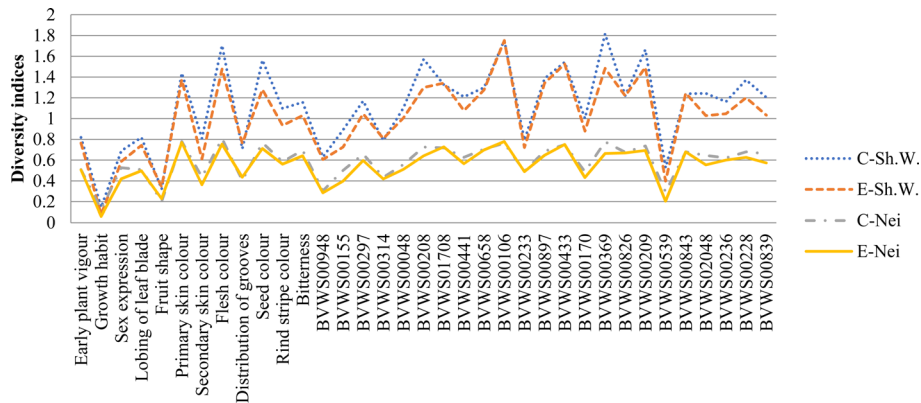
various factors, the number and source of genotypes under study and the level of polymorphism of the microsatellite markers, have the greatest impact on the magnitude of PIC. Though there are several studies on genetic diversity of watermelon, the current study includes the largest panel size only after Zhang *et al.* (2016) and the largest panel of Indian accessions ever studied.

### Population structure analysis

Insights into the genetic diversity and structure of germplasm panel are crucial for the successful deployment and efficient management of the genetic stocks (Li *et al.*, 2010). The predominance of the exotic and Indian accessions into statistically distinct subpopulations (online Supplementary Table S3) indicated that the evolution of the exotic and Indian landraces is parallel and largely unrelated to each other. However, among the Indian inbreds, the preponderance of *C. lanatus* and *C. mucospermus* into different subpopulations indicated a clear distinction between these two species, though they were reported to be taxonomically close to each other (Guo *et al.*, 2013; Renner *et al.*, 2017). Both Pop II-A and Pop II-B predominantly included *C. mucospermus* genotypes (online Supplementary Table S2), however, these populations

were statistically distinct (Table 3), indicating a genetic distinction among the accessions within the same species.

Across subpopulations, the magnitude of observed heterozygosity was less than the corresponding expected heterozygosity/gene diversity. This implies that inbreeding has happened among the individuals during the process of their evolution and/or domestication. Among the five sub-populations, Pop I-B (with 90.90% Indian accessions); comprised of *C. lanatus* (58.18%), *C. colocynthis* (40%) and *C. mucospermus* (1.82%) accessions; recorded maximum value for the average number of alleles (2.5652), maximum gene diversity (0.4184) and maximum PIC (0.6396), implying presence of greater genetic diversity among the genotypes (Table 2). However, Zhang *et al.* (2016) reported low values for gene diversity (0.30), heterozygosity (0.03) and PIC (0.26) for the East-Asian genotypes including 51 Indian accessions and found those parameters to be highest in the subpopulation dominated by the American ecotypes. Nimmakayala *et al.* (2014) grouped both East-Asian and American ecotypes into a single group while studying 130 accessions of cultivated watermelon. These minor quantitative discrepancies observed in various studies might be because of the sampling size and source of genotypes used. Overall, the current study employing a larger sample size of 213 accessions is more representative of the available variability of the region and hence recording higher diversity and PIC values in populations (Pop I-B and Pop I-C) predominated by the Indian accessions when compared to earlier studies. In the present study, the grouping of the genotypes belonging to *C. lanatus* (32) and *C. colocynthis* (22) in Pop I-B confirms their genetic closeness to each other. Earlier, Mahla *et al.* (2014) reported free gene flow between these two species resulted in the development of several intermediate forms in India. We found 19 accessions from overall, 131 from Pop I and eight from Pop II belonging to the admixture group with partial ancestry from more than one genetic background which might have resulted from the complex



**Fig. 3.** Distribution of Nei and Shannon Weaver diversity indices across phenotypic traits and molecular markers in the entire collection (E-Nei; E-Sh.W) and the core collection (C-Nei; C-Sh.W). The markers have been arranged according to their respective positions in watermelon genome.

evolutionary process involving intercrossing and gene flow between different taxa. The results indicate a probable advantage in the deployment of these adapted Indian landraces in the watermelon breeding programmes to increase the fitness value. One such example is a mega variety, Arka Manik, developed at ICAR-IIHR, Bangalore during the 1980s. It was derived from a cross between IIHR-21, which is a local collection from Jahajpur, Rajasthan, India and an American cultivar Crimson Sweet. This variety is recorded to possess resistance against downy mildew, powdery mildew and anthracnose, giving it a wide range of adaptability and is still being commercially cultivated after three decades of its release.

### Formulation of the core set

The formulated Indian core collection is comprised of 46 accessions belonging to three different species *viz.*, *C. lanatus* (29), *C. colocynthis* (14) and *C. mucospermus* (3). The size of the core set ( $n = 46/194$ ), i.e. 23.71% of the entire Indian inbred-collection used, which is slightly higher than the general size of core collections (10%) of the crop plants (Agrama *et al.*, 2009). This may be because of the collection of the accessions from the wild without much human intervention or due to a small base panel. Zhao *et al.* (2010) also concluded that accessions of wild species or landraces are richer in terms of allele frequency as compared to commercially bred varieties. Therefore, accessions belonging to different species included in our study might have widened the size of the core collection to ensure 100% allelic coverage.

Watermelon is not native to India. However, considerable diversity exists in various pockets of the country to be treated as a secondary centre of diversity. Our study signifies the importance of this diversity in the well-adapted landraces which can be used for broadening the genetic base to improve the adaptive value of the modern-day

improved varieties that otherwise have a narrow genetic base. The developed core collection would ease in maintenance of the diversity present among indigenous *Citrullus* sp. accessions; would ease trait search while exploring Indian diversity and can also be used as a base panel for association mapping. Core collections from different countries of diversity for this crop can be pooled to form a global core. Such a global core would help in understanding trait diversity in entirety aiding in successful deployment in breeding programmes.

### Supplementary material

The supplementary material for this article can be found at <https://doi.org/10.1017/S1479262120000386>.

### Acknowledgements

The first author acknowledges the University Grants Commission, Ministry of Human Resource Development, Government of India, for granting fellowship for PhD programme and the authors acknowledge Indian Council of Agricultural Research, New Delhi for funding the programme entitled 'CRP on Agrobiodiversity'.

### Author contributions

ESR, DCLR and CDP conceived and designed the experiments. MR and RNT carried out the phenotypic evaluation. SP performed the genotyping and statistical analysis with assistance from ESR and DCLR. SP and ESR wrote the manuscript. BV made necessary corrections. All the authors discussed the results and commented on the manuscript.

### Conflict of interest

The authors declare that there is no conflict of interest. The authors also declare that the manuscript has neither been submitted nor under consideration elsewhere. All authors have read and commented on the manuscript. This article does not contain any studies with human participants or animals performed by any of the authors.



## References

- Agrama HA, Yan WG, Lee F, Fjellstrom R, Chen MH, Jia M and McClung A (2009) Genetic assessment of a mini-core subset developed from the USDA Rice Gene bank. *Crop Science* 49: 1336–1346.
- Black CA (1965) Methods of Soil Analysis: Part I- Physical and Mineralogical Properties. *American Society of Agronomy*. Madison, Wisconsin, USA. 1572 p.
- Chomicki G and Renner SS (2015) Watermelon origin solved with molecular phylogenetics including linnaean material: another example of museumics. *New Phytologist* 205: 526–532.
- Dane F and Liu J (2007) Diversity and origin of cultivated and citron type watermelon (*Citrullus lanatus*). *Genetic Resources and Crop Evolution* 54: 1255–1265.
- Dane F, Liu J and Zhang C (2007) Phylogeography of the bitter apple *Citrullus colocynthis*. *Genetic Resources and Crop Evolution* 54: 327–336.
- Doyle JJ and Doyle JJ (1987) A rapid isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.
- Evanno G, Regnaut S and Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611–2620.
- Frankel OH and Brown AHD (1984) Plant genetic resources today: a critical appraisal. In: Holden JHW and Williams JT (eds.) *Crop Genetic Resources: Conservation and Evaluation*. London: Allen and Unwin Ltd. pp. 249–257.
- Guo S, Zhang J, Sun H, Salse J, Lucas WJ, Zhang H, Zheng Y, Mao L, Ren Y, Wang Z, Min J, Guo X, Murat F, Ham BK, Zhang Z, Gao S, Huang M, Xu Y, Zhong S, Bombarely A, Mueller LA, Zhao H, He H, Zhang Y, Zhang Z, Huang S, Tan T, Pang E, Lin K, Hu Q, Kuang H, Ni P, Wang B, Liu J, Kou Q, Hou W, Zou X, Jiang J, Gong G, Klee K, Schoof H, Huang Y, Hu X, Dong S, Liang D, Wang J, Wu K, Xia Y, Zhao X, Zheng Z, Xing M, Liang X, Huang B, Lv T, Wang J, Yin Y, Yi H, Li R, Wu M, Levi A, Zhang X, Giovannoni JJ, Wang J, Li Y, Fei Z and Xu Y (2013) The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nature Genetics* 45: 51–58.
- Hwang J, Kang J, Son B, Kim K and Park Y (2011) Genetic diversity in watermelon cultivars and related species based on AFLP and EST-ISSRs. *Notulae Botanicae Horti Agrobotanici* 39: 285–292.
- Jakobsson M and Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics (Oxford, England)* 23: 1801–1806.
- Kim KW, Chung HK, Cho GT, Ma KH, Chandrabalan D, Gwag JG, Kim TS, Cho EG and Park YJ (2007) Powercore: a program applying the advanced M strategy with a heuristic search for establishing allele mining sets. *Bioinformatics (Oxford, England)* 23: 2155–2162.
- Kwon YS, Oh YH, Yi SI, Kim HY, An JM, Yang SG, Ok SH and Shin JS (2010) Informative SSR markers for commercial variety discrimination in watermelon (*Citrullus lanatus*). *Genes and Genomics* 32: 115–122.
- Kyriacou MC, Roupheal Y, Colla G, Zrenner R and Schwarz D (2017) Vegetable grafting: the implications of a growing agronomic imperative for vegetable fruit quality and nutritive value. *Frontiers in Plant Science* 8: 741. <https://www.frontiersin.org/article/10.3389/fpls.2017.00741>.
- Levi A, Thies JA, Wechter WP, Harrison HF, Simmons AM, Reddy UK, Nimmakayala P and Fei Z (2013) High frequency oligonucleotides: targeting active gene (HFO-TAG) markers revealed wide genetic diversity among *Citrullus* Spp. accessions useful for enhancing disease or pest resistance in watermelon cultivars. *Genet Resources and Crop Evolution* 60: 427–440.
- Levi A, Jarret R, Kousik S, Wechter WP, Nimmakayala P, Reddy U (2017) Genetic resources of watermelon. In: Grumet R, Katzir N and Garcia-Mas J (eds.) *Genetics and Genomics of Cucurbitaceae, Plant Genetics and Genomics: Crops and Models*. Cham, Switzerland: Springer International Publishing AG, pp. 87–110.
- Li J, Schulz B and Stich B (2010) Population structure and genetic diversity in elite sugar beet germplasm investigated with SSR markers. *Euphytica* 175: 35–42.
- Liu KJ and Muse SV (2005) Powermarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics (Oxford, England)* 21: 2128–2129.
- Mahla HR, Singh JP and Roy MM (2014) *Seed Purpose Watermelon in Arid Zone*. Jodhpur, Rajasthan, India: Central Arid Zone Research Institute, pp. 1–39.
- Minsart LA, Zoro IA, Dje Y, Baudoin JP, Jacquemart AL and Bertin P (2011) Set up of simple sequence repeat markers and first investigation of the genetic diversity of West African watermelon (*Citrullus lanatus* ssp. *vulgaris* Oleaginous type). *Genetic Resources and Crop Evolution* 58: 805–814.
- Mujaju C, Zborowska A, Werlemark G, Garkava-Gustavsson L, Andersen SB and Nybom H (2011) Genetic diversity among and within watermelon (*Citrullus lanatus*) landraces in Southern Africa. *Journal of Horticultural Science and Biotechnology* 86: 353–358.
- Munisse P, Jensen BD and Andersen SB (2013) Genetic differentiation of watermelon landraces in Mozambique using microsatellite markers. *African Journal of Biotechnology* 12: 5513–5521.
- Nantoume AD, Andersen SB and Jensen BD (2013) Genetic differentiation of watermelon landrace types in Mali revealed by microsatellite (SSR) markers. *Genetic Resources and Crop Evolution* 60: 2129–2141.
- Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America* 70: 3321–3323.
- Nimmakayala P, Abburi VL, Bhandary A, Abburi L, Vajja VG, Reddy R, Malkaram S, Venkatramana P, Wijeratne A, Tomason YR, Levi A, Wehner TC and Reddy UK (2014) Use of VeraCode 384-plex assays for watermelon diversity analysis and integrated genetic map of watermelon with single nucleotide polymorphisms and simple sequence repeats. *Molecular Breeding* 34: 537–548.
- Olsen SR, Cole CV, Watanabe FS and Dean LA (1954) *Estimation of Available Phosphorus in Soils by Extraction with Sodium Bicarbonate*. Washington: Circular of US Department of Agriculture. 939, 19.
- Paris HS, Daunay MC and Janick J (2013) Medieval iconography of watermelons in Mediterranean Europe. *Annals of Botany* 112: 867–879.
- Perrier X and Jacquemoud-Collet JP (2006) DARwin software. Available at: <http://darwin.cirad.fr/>.
- Pritchard JK, Stephens M and Donnelly P (2000) Inference of population structure from multilocus genotype data. *Genetics* 155: 945–959.
- Rao ES, Kadirvel P, Symonds RC, Geethanjali S and Ebert AW (2012) Using SSR markers to map genetic diversity and population structure of *Solanum pimpinellifolium* for development of a core collection. *Plant Genet Resources* 10: 38–48.
- Reddy UK, Abburi L, Abburi VL, Saminathan T, Cantrell R, Vajja VG, Reddy R, Tomason YR, Levi A, Wehner TC and Nimmakayala

- P (2015) A genome-wide scan of selective sweeps and association mapping of fruit traits using microsatellite markers in watermelon. *Journal of Heredity* 106: 166–176.
- Ren Y, Zhao H, Kou Q, Jiang J, Guo S, Zhang H, Hou W, Zou X, Sun H, Gong G, Levi A and Xu Y (2012) A high resolution genetic map anchoring scaffolds of the sequenced watermelon genome. *PLoS One* 7: e29453.
- Renner SS, Sousa A and Chomicki G (2017) Chromosome numbers, Sudanese wild forms, and classification of the watermelon genus *Citrullus*, with 50 names allocated to seven biological species. *Taxonomy* 66: 1393–1405.
- Shannon CE and Weaver W (1963) *The Mathematical Theory of Communication*. Urbana: University of Illinois Press. (First published in 1949).
- Singh D, Singh R, Sandhu JS and Chunneja P (2017) Morphological and genetic diversity analysis *Citrullus* Landraces from India and their genetic interrelationship with continental watermelons. *Scientia Horticulturae* 218: 240–248.
- Subbiah BV and Asija GL (1956) A rapid procedure for the estimation of available nitrogen in soils. *Current Science* 25: 259–260.
- Wehner TC (2008) Watermelon. In: Prohens J and Nuez F (eds.). *Handbook of Plant Breeding: Vegetables I: Asteraceae, Brassicaceae, Chenopodiaceae, and Cucurbitaceae*. New York: Springer Science Business LLC, pp. 381–418.
- Wei C, Chen X, Wang Z, Liu Q, Li H, Zhang Y, Ma J, Yang J and Zhang X (2017) Genetic mapping of the Lobed Leaf 1 (*CILL1*) gene to a 127.6-kb region in watermelon (*Citrullus lanatus* L.). *PLoS One* 12: e0180741.
- Zhang H, Wang H, Guo S, Ren Y, Gong G, Weng Y and Xu Y (2012) Identification and validation of a core set of microsatellite markers for genetic diversity analysis in watermelon, *Citrullus lanatus* thumb. Matsum. Nakai. *Euphytica* 186: 329–342.
- Zhang H, Fan J, Guo S, Ren Y, Gong G, Zhang J, Weng Y, Davis A and Xu Y (2016) Genetic diversity, population structure and formation of core collection of 1197 *Citrullus* Accessions. *HortScience* 51: 23–29.
- Zhao R (2015) *A History of Food Culture in China*. New York: SCPG Publishing Co.
- Zhao W, Cho GT, Ma KH, Chung JW, Gwag JG and Park YJ (2010) Development of an allele mining set in rice using a heuristic algorithm and SSR genotype data with least redundancy for the post-genomic era. *Molecular Breeding* 26: 639–651.
- Zoltán T, Gyulai G, Szabó Z, Horváth L and Heszky L (2007) Watermelon (*Citrullus lanatus*) production in Hungary from the Middle Ages (13th century). *Hungarian Agricultural Research* 4: 14–19.