J. Hortic. Sci.
Vol. 18(1) : 46-52, 2023

https://doi.org/10.24154/jhs.v18i1.2146

Original Research Paper

SSR analysis to assess genetic diversity and population structure in parthenocarpy cucumber (*Cucumis sativus* L.)

Kaur M.*, Sharma P., Sharma A., Hemalata and Kumar N.

Department of Vegetable Science and Floriculture, College of Agriculture, CSK HPKV, Palampur - 176062 (HP) India

*Correspondence author Email : sidhumanpreetk12@gmail.com

ABSTRACT

The genetic diversity and population relationship was determined in 14 genotypes of parthenocarpic cucumber (*Cucumis sativus* L.) using simple sequence repeats (SSR) markers. In this study, fifty-nine SSR markers comprehensively showed polymorphism among cucumber genotypes. Total 252 alleles were identified with an average of 4.27 alleles per locus, while the polymorphism information content (PIC) of the primers ranged from 0.34 to 0.84 with a mean value of 0.62. The major allele frequency and heterozygosity ranged from 0.21 to 0.75 and from 0.43 to 0.89, respectively. Maximum major allele frequency was reported with primer Cs-Female-4, whereas the maximum value of polymorphic information content was found with the primer SSR11742. The dendrogram clustered genotypes into two main groups A and B with 8 and 6 genotypes, respectively. Jaccard's similarity coefficient ranged from 0.63 to 0.86 with maximum similarity between genotypes DDPCG3 and PLP-1, whereas minimum similarity was observed between DDPCG8 and PLP Gy-1-08B. The population structure revealed three sub-populations with some admixtures. Principal coordinate analysis (PCoA) with SSR markers revealed that the genotypes were uniformly distributed across the two axes in both the plots with 41.76% of cumulative variation. The genetic divergence within indigenous genotypes allow genotypic identification, gene mapping and cloning for improvement in cucumber breeding.

Keywords : Cucumber, Genetic diversity, Polymorphism, Population structure, SSR Markers

INTRODUCTION

Cucumber is a member of the diverse and distinct Cucurbitaceae family and is widely grown for both fresh and processing purposes around the world. Primary centre of origin was India where both wild and cultivated species exist while, China and near east are secondary centre of origin (Telford and Renner, 2010). Both cultivated and wild species viz., *Cucumis sativus* var. *hardwickii* render enormous variation for various traits like growth habit, sex expression, fruit size, spines and flesh bitterness. About 70% of the cucumber world production is contributed by Asian countries, Turkey, Iran and Russia. In India, cucumber covered an area of 104 thousand hectares with 1603 thousand MT annual production (NHB 2019).

Cucumber is an ideal model crop for genetic studies due to smaller genome size of approximately 367Mb with shorter life cycle (Kaur and Sharma, 2021). Breeding cucumber for enhancing yield, quality, and biotic and abiotic stress tolerance is a major challenge for the breeders, globally (Yuan *et al*., 2008). In spite of huge variability, it has narrow genetic base with only 12% polymorphism which limits the new cultivar development by cross breeding (Pandey *et al.,* 2018). There is scope for improvement of the productivity with the use of improved varieties or hybrids of cucumber (Pandey *et al.,* 2016). Selection of suitable parents for breeding programme depends on the existence of variability in the germplasm. Identification of the suitable parents is the most imperative for hybridization. Recent progresses in plant genomic offers an opportunity for assessing genetic diversity through use of molecular markers (Yang *et al.,* 2015). Molecular markers are more advantageous than morphological characters due to more stability under variable environment conditions. Different types of molecular markers are random amplified polymorphic DNA (RAPD), sequence characterized amplified regions (SCAR), amplified fragment length polymorphisms (AFLP) and simple sequence repeats (SSR) (Dar *et al.,* 2017). Among all, SSR markers are widely used in plant genomics like gene mapping, quantitative trait loci (QTL), marker assisted selection (MAS), evolutionary studies and genetic diversity

analysis (Mahajan *et al.,* 2016). SSR markers are used in cucumber for assessment of genetic diversity in cucumber (Yang *et al.,* 2015). Genetic diversity and population structure is very important for the maintenance, conservation and improvement in productivity in agriculture. Plant genetic diversity can be preserved and stored in the form of plant genetic resources in gene banks and DNA libraries for long term conservation. These plant genetic resources could be utilized in future for the crop improvement against various biotic and abiotic stresses to meet global food security (Garzon-Martinez *et al.,* 2015). Due to narrow genetic base and use of limited number of SSR markers for genetic diversity analysis, there is a dire need for studying genetic diversity using SSR markers for bridging the gap in the crop improvement by hybridization. Therefore, this study was focused to determine genetic diversity and population structure using SSR markers in cucumber. The findings of this work will aid in the selection of cucumber genotypes with a high genetic diversity of the genes used in crossbreeding, QTL mapping, gene tagging and other imperative genomic studies

MATERIALS AND METHODS

Experimental material

This study was conducted at Research Farm of Vegetable Science, Department of Vegetable Science and Floriculture (N 32° 6[1], E 76° 3[1]), CSK-HPKV, Palampur. Agro-climatically, it is located in the mid-hill regions having humid sub-temperate climate with 2,500 mm annual rainfall. The experiment material comprised of fourteen genotypes both gynoecious parthenocarpic which were collected from CSK-HPKV (Palampur), PAU (Ludhiana, Punjab) and GBPUA&T (Pant Nagar) (Table 1). The genotypes were maintained at Experimental Farm and Molecular Biology Laboratory of Vegetable Science and Floriculture department, CSK- Himachal Pradesh Krishi Vishvavidyalaya, Palampur, India during the year 2020-21 to take up genetic diversity analysis.

Genomic DNA extraction and PCR amplification using SSR markers

About 5 g of plant tissue was finely ground in liquid nitrogen. The entire genomic DNA was extracted from each genotype using the CTAB technique (Doyle and Doyle, 1987). The DNA quantification was done using Nanodrop spectrophotometer at the OD 260/280 value and 0.8% agarose Gel-electrophoresis. For PCR, DNA was diluted to 50 ng/ul and refrigerated at 4°C, whereas concentrated DNA stocks were kept at -80°C for later use.

Table 1 : Cucumber germplasm and their sources used for diversity analysis

Germplasm	Collection Source
DDPCG4	CSKHPKV, Palampur
HPK-1	CSKHPKV, Palampur
Punjab Kheera-1 (PK-1)	PAU, Ludhiana
PPC-2	GBPUA&T, Pant Nagar
PPC-3	GBPUA&T, Pant Nagar
DDPCW1	CSKHPKV, Palampur
DDPCG2	CSKHPKV, Palampur
DDPCG5	CSKHPKV, Palampur
DDPCG6	CSKHPKV, Palampur
DDPCG7	CSKHPKV, Palampur
PLPGy-1-08-A (green)	CSKHPKV, Palampur
PLP Gy-1-08-B (white)	CSKHPKV, Palampur
DDPCG3	CSKHPKV, Palampur
PLP-1	CSKHPKV, Palampur

For amplification of genomic DNA, a reaction mixture of 15 µl volume was prepared using template DNA (50 ng/µl), forward and reverse primer (5µM each), MgCl2 (1.6 mM), 1 X PCR buffer (1 X: 10mM Tris-HCl, 50mM KCl, pH 8.3), dNTP mix (0.25 mM) and Taq polymerase (0.75 U/µl). The PCR reaction was carried out in thermal cycler with initial denaturation at 94°C at 3-5 min, 35-36 cycles of denaturation of 94°C for 30- 60 sec, Annealing of 50-60°C for 30-60 sec, extension of 72°C for 60-80 sec and followed by final extension of 72°C for 5-10 min. The amplified products were resolved in 3 per cent agarose gel with 100 bp ladder and gels were visualized using the gel-documentation unit (Bio-Rad).

Statistical analysis

For all analyzed genotypes, exclusive DNA bands were evaluated as present (1) or absent (0). In the SIMQUAL programme of the NTSYSpc package (version 2.02), the binary data were used to generate a Jaccard's similarity coefficient through UPGMA (unweighted pair-group method with arithmetic averages) method which allowed to design a dendrogram by genotype clustering. PIC value calculates the informativeness of a particular DNA marker (Spooner *et al.,* 1993). Using the software STRUCTURE version 2.3.4 (Pritchard *et al.,* 2000), model-based cluster analysis was performed to

determine the genetic structure and number of clusters in the data set. The number of hypothesized populations (K) varied between 2 and 10 and the analysis was carried out twice and the true k was determined according to the method described by Evanno *et al.* (2005). The run with maximum likelihood was used to assign individual genotypes into groups. POPGENE was used to calculate a variety of genetic variation parameters. Using the DARwin software version 5.0, a Neighbor-Joining tree (UnWeighted) was constructed from the dissimilarity matrix (Perrier and Jacquemoud, 2006). 1000 bootstraps were used to test branch robustness. Principal coordinates analysis (PCoA) in GenALEx 6.5 was used to visualize the genetic relationship patterns in the matrix. Structure analysis was done to estimate population structure (Q matrix) using STRUCTURE (Pritchard *et al.,* 2000; Falush *et al.,* 2003) and express as membership probability. To estimate the actual population substructure, ten different Ks (from K=1 to K=10, where K is the kinship matrix) were utilized.

RESULTS AND DISCUSSION

SSR and marker informativeness

The gel electrophoresis results for 14 germplasm with primer SSR11742 is presented in Fig. 1. The total molecular variability parameters such as PIC, heterozygosity, major allele frequency, number of alleles and allele size across all 14 germplasm are presented in Table 2 (Supplimentary file). Out of 61 SSR primers, 59 primers exhibited polymorphism. A total of 252 amplicons were created, with sizes ranging from 100 to 380 bp. The total number of alleles from 59 primers observed was 252 with a mean of 4.27 alleles per locus and eight alleles were identified in SSR11742 and SSR04689. Major allele frequency varied from 0.21 (SSR04689) to 0.75 (Cs-Female-4) with an average value of 0.42. The polymorphic information content (PIC), ranged from 0.34 (SSR30647) to 0.84 (SSR11742), with an average value of 0.62 per primer. Similarly, heterozygosity varied from 0.43 (Cs-Female-4) to 0.89 (SSR 11742) with an average value of 0.70.

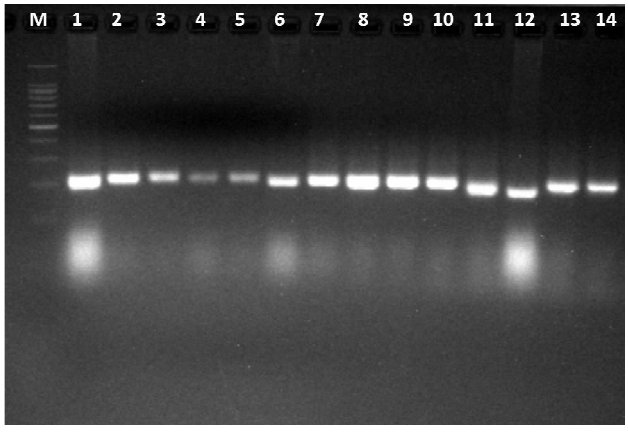


Fig. 1 : DNA profile of 14 germplasm of cucumber showing polymorphism with primer SSR11742 (M-100 bp ladder)

Genetic diversity assessment and structure analysis

Fourteen cucumber genotypes were divided into two main clusters (A and B). Cluster A was split into two sub-clusters comprising of total of 8 germplasm, while Cluster B had contained six genotypes namely, DDPCG6, DDPCG7, PLPGy-1-08A, PLPGy-1-08B, DDPCG3 and PLP-1 (Fig. 2). Based on UPGMA analysis, Jaccard's similarity coefficient varied from 0.63 to 0.84 with maximum similarity between genotype DDPCG3 and PLP-1 (0.86), whereas minimum similarity was between DDPCG8 and PLPGy-1-08B (0.59). Based on Neighbor Joining analysis, genotypes were grouped into three clusters as depicted using the color codes in Fig. 3. Cluster I (Red), Cluster II (Blue) and Cluster III (Green)

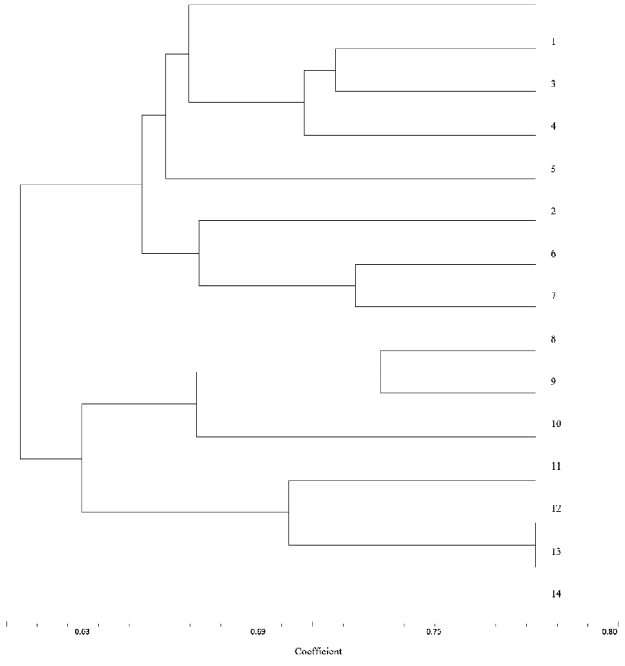


Fig. 2 : Dendrogram depicting genetic relationships among the cucumber germplasm constructed by NTSYS–PC (version 2.02) using UPGMA method

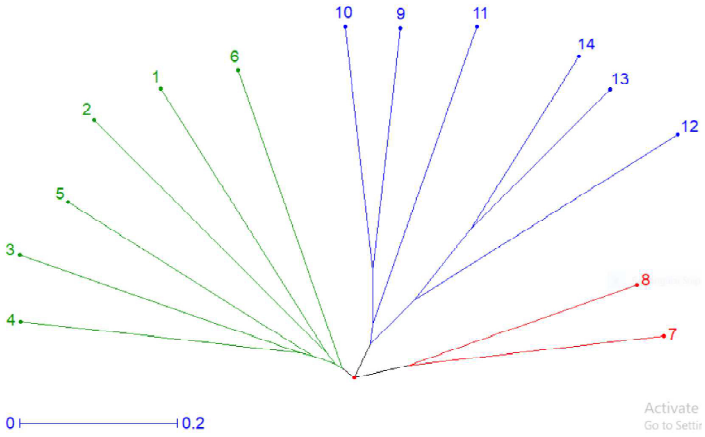
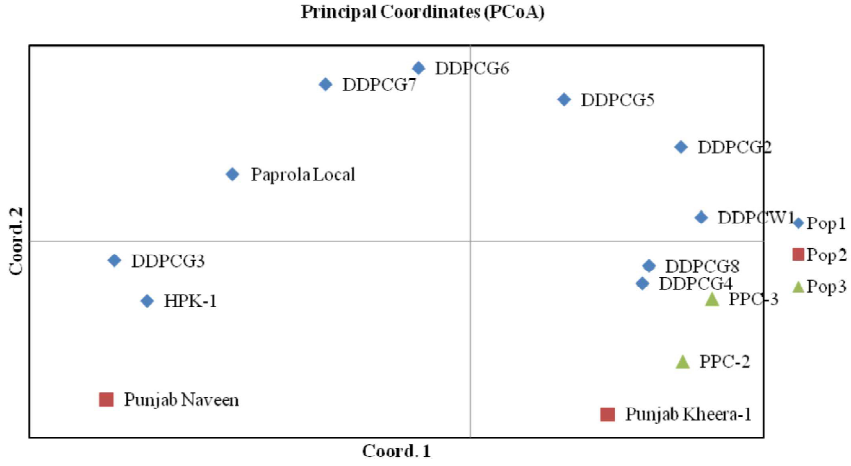


Fig. 3 : Neighbor-Joining tree of cucumber germplasm using SSR markers generated by DARwin software



Percentage of variation explained by the first 3 axes

Axis	1	2	3
%	17.59	14.10	10.08
Cum %	17.59	31.68	41.76

Fig. 4 : PCoA scatter diagram analysis showing the distribution of 14 cucumber germplasm

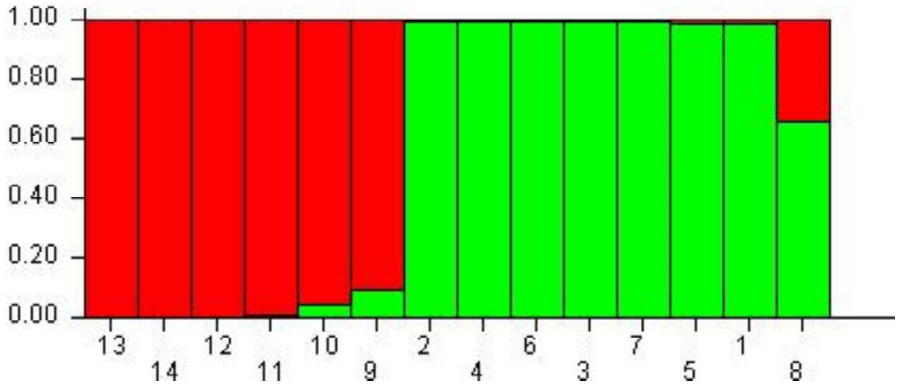


Fig. 5 : Genetic structure of 14 cucumber germplasm (red and green) represent the two groups, defined by the K value. Cucumber germplasm showing more than one color may have an admixture

comprised of two, six and six genotypes, respectively. Principal coordinate analysis (PCoA) showed that first three coordinates accounted for 41.76% cumulative variation among 14 genotypes (Fig. 4) with the first and second coordinates explaining 17.59% and 14.10% of the total variation respectively.

The STRUCTURE analysis divided the population into two groups. The differentiations at K =2 were nearly equivalent to pedigree knowledge with a few outliers. In group 1 (Red) consists of 6 genotypes and group 2 (Green) comprises 8 genotypes (Fig. 5). The germplasm generated by the NTSYS software were confirmed using STRUCTURE analysis at K = 2. As a result of this, it was established that the germplasm that were separated according to cluster analysis were almost identical to those that were divided according to structure analysis, with a few minor differences.

The genetic diversity and population structure in cucumber was investigated for improvement of various traits using crop breeding practices. A limited number of SSR molecular markers were used with Indian cucumber genotypes. It has been observed that SSR markers showed high polymorphism in cucumber. In our study, we have determined the genetic diversity using sixty-one SSR markers in 14 genotypes of cucumber comprising a wider geographical distribution of genotypes. Among 61 SSRs primers, 59 primers showed high polymorphism and a total of 252 alleles were identified with an amplicon size ranging from 100-380 bp. The number of alleles varied from 2-8 with a mean of 4.27 alleles per locus. Similarly, Dar *et al.* (2017) and Lv *et al.* (2012) observed an average number of alleles 2.9 and 13.7 per locus, respectively. The polymorphic information content (PIC), a measure related to marker discrimination, ranged from 0.34 (SSR30647) to 0.84 (SSR11742), with a mean of 0.62 per primer. Our study revealed similar results of PIC (0.62) in comparison with previous reports on cucumber *i.e.,* 0.664 and 0.69 (Hu *et al*., 2011; Normohamadi *et al.,* 2017) while, PIC was lower in Indian cucumber (0.310), Chinese cucumber (0.388) and cucumber (0.33) (Hu *et al.,* 2011; Pandey *et al.,* 2013; Dar *et al.,* 2017). A range of 0.12-0.44 was observed for PIC value for 15 primers with the mean value of 0.21 (Someh *et al.,* 2016). SSR11742 and SSR04689 markers were found more polymorphic among 59 SSR markers due to their high PIC values. The results were in agreement with earlier studies on cucumber suggesting the role of SSR markers for identification of genotypes, DNA fingerprinting and maintenance of genotypes in the gene banks. Based on UPGMA analysis with Jaccard's similarity coefficient varied from 0.63 to 0.86. Similarly, Someh *et al.* (2016) and Normohamadi *et al*. (2017) reported Jaccard's similarity coefficient ranging from 0.56 to 0.88 and 0.51 to 0.92 in cucumber, respectively. Lower range of Jaccard's similarity coefficient *viz*., 0.01-0.44 and 0.35-0.51 was reported in cucumber by Valcarcel *et al*. (2018) and Park *et al*. (2021). There was no regional distribution trend in the clustering pattern based on UPGMA and PCA. This could be due to regular gene flow through seed exchange between different places, which is most likely due to human interference (Garzon-Martinez *et al.,* 2015). Minimum Jaccard's similarity coefficient was observed in DDPCG8 and PLPGy-1-08B showing maximum diversity among genotypes. The genotypes DDPCG8 and PLPGy-1-08B were collected from different parts of Indian origin The clustering formed by the UPGMA dendrogram was moderately validated by projecting individual genotypes into a two-dimensional multivariate space in PCoA diagram. As per UPGMA method the cucumber genotypes were divided into two main clusters A (A_1-5 and A_2-3) and B (6). Similar results were reported by Dar *et al.* (2017) which grouped cucumber germplasm into two main distinct clusters. Various clustering methods were employed to assess genetic relationship of different genotypes or germplasm. Based on Neighbour Joining, fourteen genotypes were grouped into three clusters as represented by using color codes. Cluster I consists of 2 genotypes followed by 6 genotypes in cluster II and III.

PCoA is a multivariate strategy for grouping data based on similarity coefficients or variance or covariance values that provides more information about main groups, whereas cluster analysis provides higher resolution among closely related populations. PCoA explores correlations between many quantitative variables by constructing a small number of linear combinations (principal components) that retain as much information as feasible from the original data. Principal coordinate analysis (PCoA) showed that first three coordinates accounted for 41.76% cumulative variation among 14 genotypes with the first and second coordinates explaining 17.59% and 14.10% of the

total variation respectively. The population structure analysis grouped the genotypes into 2 groups including genotypes having admixtures. As a result, pedigree information was combined with cluster membership to determine the division of Red and Green groupings. Similar results were reported in cucumber (Pandey et al. 2013; Dar *et al.,* 2017) and Turkish melons (Sensoy *et al.,* 2007). The increased variance should be recorded for germplasm preservation and agricultural enhancement breeding strategies.

CONCLUSION

This study could be used to estimate genetic variation within a group of elite genotypes to employ in cucumber improvement in India. A total of 14 cucumber genotypes were assessed using 59 polymorphic SSR markers. The experiment depicted total number of 252 amplicons, with an overall average of 4.27 alleles per locus. SSR 11742 primer was recorded to have good marker informativeness. Based on UPGMA cluster analysis, maximum similarity (less diverse) was observed between genotype DDPCG3 and PLP-1 whereas minimum similarity (more diverse) between DDPCG8 and PLPGy-1-08B. The population structure depicted three main populations including admixture genotypes. It may be further utilized in future projects related to QTLs identification, genome wide association studies, DNA fingerprinting and preservation of cucumber germplasm across India and other countries.

REFERENCES

Dar, A. A., Mahajan, R., Lay, P. and Sharma, S. 2017. Genetic diversity and population structure of *Cucumis sativus* L. by using SSR markers. *3 Biotech.,* **7**: 307.

Doyle, J. J. and Doyle, J. L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull*., **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. *Mol. Ecol*., **14**: 2611-2620.

Falush, D., Stephens, M. and Pritchard, J. K. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics,* **164**: 1567-1587.

Garzon-Martinez, G. A., Osorio-Guarin, J. A., Delgadillo-Duran, P., Mayorga, F., Enciso-Rodriguez, F. E., Landsman, D., Marino-Ramirez, L. and Barrero, L. S. 2015. Genetic diversity and population structure in *Physalis peruviana* and related taxa based on InDels and SNPs derived from COSII and IRG markers. *Plant Gene,* **4**: 29-37.

Hu, J., Wang, L. and Li, J. 2011. Comparison of genomic SSR and EST-SSR markers for estimating genetic diversity in cucumber. *Biol. Plant*., **55**: 577-580.

Kaur, M. and Sharma, P. 2021. Recent advances in cucumber (*Cucumis sativus* L.). *J. Hortic. Sci. Biotechnol.,* doi: 10.1080/14620316.2021.1945956

Lv, J., Qi, J., Shi, Q., Shen, D., Zhang, S., Treuren, R., Dooijeweert, W. and Zhang, Z. 2012. Genetic diversity and population structure of cucumber (*Cucumis sativus* L.). *PLoS One,* **7**: e46919.

Mahajan, R., Zargar, S. M., Singh, R., Salgotra, R. K., Farhat, S. and Sonah, H. 2016. Population structure analysis and selection of core set among common bean genotypes from Jammu and Kashmir., India. *Appl. Biochem.,* **182**: 16-28.

Normohamadi, S., Solouki, M. and Heidari, F. 2017. Diversity in cucumber genotypes based on morphological traits and SSR molecular markers. *Biosci Biotech. Res. Asia* **14**: 775-782.

Pandey, S., Ansari, W.A., Atri, N., Singh, B., Gupta, S. and Bhat, K.V. 2016. Standardization of screening technique and evaluation of muskmelon genotypes for drought tolerance. *Plant Genet Resour Charact Util*. https://doi.org/10.1017/S1479262116000253

Pandey, S., Ansari, W.A., Mishra, V. K., Singh, A.K. and Singh, M. 2013. Genetic diversity in Indian cucumber based on microsatellite and morphological markers. *Biochem. Syst. Ecol.,* **51**: 19-27.

Pandey, S., Ansari, W. A., Pandey, M. and Singh, B. 2017. Genetic diversity of cucumber estimated by morpho-physiological and EST-SSR markers. *Physiol. Mol. Biol. Plants*.

Pandey, S., Ansari, W.A., Pandey, M. and Singh, B., 2018. Genetic diversity of cucumber estimated by morpho-physiological and EST-SSR markers. *Physiol. Mol. Biol. Plants*, **24**: 135-146.

Park, G., Choi, Y., Jung, J. K., Shim, E. J., Kang, M. Y., Sim, S. C., Chung, S. M., Lee, G. P. and Park, Y. 2021. Genetic diversity assessment and cultivar identification of cucumber (C*ucumis sativus* L.) using the fluidigm single nucleotide polymorphism assay. *Plants,* **10**: 395.

Perrier, X. and Jacquemoud, C. 2006. DARwin software.

Pritchard, J. K., Stephens, M. and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics*, **155**: 945-959.

Sensoy, S., Buyukalaca, S. and Abak, K. 2007. Evaluation of genetic diversity in Turkish melons (*Cucumis melo* L.) based on phenotypic characters and RAPD markers. *Genet. Resour. Crop Evol*., **54**: 1351-1365.

Someh, M., Kiani, G., Ranjbar, G. A. and Alavi, S. M. 2016. Assessment of genetic diversity in cucumber varieties using RAPD markers. *J. Appl. Hortic*., **18**: 64-67.

Spooner, D. M., Anderson, G. J. and Jansen, R. K. 1993. Chloroplast DNA evidence for the inter-relationships of tomatoes, potatoes, and pepinos (Solanaceae). *Am. J. Bot*., **80**: 676.

Valcarcel, J. V., Perez-de-Castro, A., Diez, M. J. and Peiro, R. 2018. Molecular characterization of the cucumber (*Cucumis sativus* L.) accessions held at the COMAV's genebank. *Spanish J. Agric. Res*., **16**: 2171-9292.

Yang, Y. T., Liu, Y., Qi, F., Xu, L. L. and Li, X. Z. 2015. Assessment of genetic diversity of cucumber cultivars in China based on simple sequence repeats and fruit traits. *Genet. Mol. Res.,* **14**: 19028-19039

Yuan, X. J., Pan, J. S., Cai, R., Guan, Y., Liu, L. Z., Zhang, W. W., Li, Z., He, H. L., Zhang, C., Si, L. T. and Zhu, L. H. 2008. Genetic mapping and QTL analysis of fruit and flower related traits in cucumber (*Cucumis sativus* L.) using recombinant inbred lines. *Euphytica,* **164**: 473-491.

(Received : 22.09.2022; Revised : 10.12.2022; Accepted 12.12.2022)