Genome-wide SSR Markers in Bottle Gourd: Development, Characterization, Utilization in Assessment of Genetic Diversity of National Genebank of India and Synteny With Other Related Cucurbits

Bhawna Bonthala

NBPGR

Manjusha Verma (✉ manjusha.verma@icar.gov.in)

National Bureau of Plant Genetic Resources <https://orcid.org/0000-0001-7711-5242>

M Z Abdin

Jamia Hamdard University

Lalit Arya

NBPGR

Chithra D Pandey

NBPGR

Ved Dixit

NBPGR

Priya Yadav

NBPGR

Research Article

Keywords: Bottle gourd, Microsatellites, Genome-wide, SSR marker development, RAD-Seq, Synteny

Posted Date: November 19th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-990222/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at Journal of Applied Genetics on February 1st, 2022. See the published version at <https://doi.org/10.1007/s13353-022-00684-1>.

Abstract

Lagenaria siceraria (Molina) Standley is an important cultivated crop with its immense importance in pharmaceutical industry and as vegetable. Its seed, root, stem, leaves, flower and fruit are used as an ointment for ailment of various diseases throughout Asia. Despite its worldwide importance, informative co-dominant microsatellite markers in the bottle gourd crop are very restricted, impeding genetic improvement, cultivar identification and phylogenetic studies. Next generation sequencing has revolutionized the approaches for discovery, assessment and validation of molecular markers. We conducted a genome wide analysis, for developing SSR markers by utilizing restriction site-associated DNA sequencing (RAD-Seq) data. By performing *in silico* mining of microsatellite repeat-motifs, we developed 45,066 perfect SSR markers. Of which 203 markers were successfully validated and 101 (49.75%) polymorphic primer pairs were utilized for an in depth genetic diversity and population structure analysis of 96 accessions from the National Genebank of India. Tetranucleotide repeats (~34.3%) were the most prevalent followed by trinucleotide repeats (~30.73%), further 21.03%, 9.6% and 4.3% of di-, penta- and hexa- nucleotide repeats in the bottle gourd genome. Synteny of SSR markers on 11 bottle gourd linkage groups was correlated with the 7 chromosomes of cucumber (93.2%), 12 chromosomes of melon (87.4%) and 11 watermelon (90.8%). The generated SSR markers provide a valuable tool for germplasm characterization, genetic linkage map construction, studying synteny, gene discovery and for breeding in bottle gourd and other cucurbits species.

Key Message

Development of 45,066 perfect microsatellite markers as a valuable tool for marker assisted selection (MAS) in plant breeding.

1. Introduction

Bottle gourd [*Lagenaria siceraria* (Mol.) Standl.] ($2n = 2x = 22$), is an edible, medicinal, container and a grafting stock plant cultivated all over the tropics (Heise et al., 1979). The bottle gourd is known as one of the first crops to be cultivated (>10,000 years ago) (Whitaker et al., 1971; Erickson et al., 2005) worldwide indicating its importance in agro industry. High genetic variability exists in bottle gourd, especially in fruit size and shape (Emina et al., 2012). It's reported that bottle gourd is loaded with various medical properties like anti-

cancerous, cardio-protective (Fard et al., 2008), diuretic, purgative, cooling effects (Badmanaban et al., 2010). It can also cure ulcers, pectoral cough, asthma and other bronchial disorders (Upaganlawar et al., 2010). Its seed and fruit extracts (Prashant et al., 2014; Nidhi et al., 2017) exhibit a variety of pharmacological properties and are utilized in the pharmaceutical industry (Sakshi et al., 2015) as well as Ayurveda (Nidhi et al., 2017). Bottle gourd fruit is used in craft industries as a raw material for the production of artifacts, musical instruments, and containers for food and beverages (Danieli et al., 2020). The commercial potential of the crop increases by understanding its fruit juice potential in reducing weight, to cure baldness and aids in preventing tooth decay (Nidhi et al., 2017). Bottle gourd seed oil is used in cosmetic industry for skin therapy, in cosmetics and in treatment of hyperplasia (Prashant et al., 2014). Many technologies are being developed to use bottle gourd in many forms, such as making bottle gourd ice cream (Barot et al., 2014) and adding fruit like Jamun to increase the shelf life of bottle gourd beverages (Palamthodi et al., 2019). Though there is an increased demand for cultivation of bottle gourd due to its immense cultivation, craft and medicinal importance, but very scanty information is available about the genetic diversity of its germplasm.

Among various available molecular markers, microsatellites markers are ideal due to their ease of use and co-dominance. Microsatellites or simple sequence repeats (SSRs) are small (1-6 bp) tandem repeats, which are ubiquitously found in the genome of both prokaryotic and eukaryotic organisms (Toth et al., 2000). High mutation rate, are the consequences of the change in array length of allele, these repeats signify a rich cause of hyper-variable co-dominant markers (Morgante et al., 2002; Powell et al., 1996). Consequently, microsatellite markers are considered as an important markers in various research such as population genetics (Innan et al., 1997), linkage mapping (McCouch et al., 2002; Somers et al., 2004), phylogenetics, structural, functional or comparative genomics research (Garza et al., 1995; MacHugh et al., 1997). Due to their high reproducibility, multi-allelic variation, co-dominant inheritance and abundance in the genome (Tautz et al., 1984), they are also widely used for marker-assisted breeding and parentage analysis (Bowers et al., 1997). Microsatellites are believed to play significant roles in genome evolution, by producing genetic variability (Kashi et al., 1997), study of gene expression regulation (Santi et al., 2003; Saveliev et al., 2003), acclimatizing the development of the cell cycle (Li et al., 2002), etc. As shown in different SSR data from plants (Morgante et

al., 2002), humans (Subramanian et al., 2003), and other eukaryotic species (Toth et al., 2000), microsatellites are distributed around the genome in an uneven and non-random manner, with both similarities and differences observed among taxa. The distribution of microsatellites in the genome provides important information on their utility as molecular markers. Genomic SSR markers are more transferable among related species, making them easier to employ as anchor markers for comparative physical mapping (Varshney et al., 2005).

Technological advancements such as genotyping-by-sequencing (GBS) or reduced-representation libraries sequencing (RRLS) appeared to be the most promising technical replacements (Altshuler et al., 2000; Elshire et al., 2011; Helyar et al., 2011). Most significantly, restriction site-associated DNA sequencing (RAD-Seq), one of the techniques known as a genome 'complexity reduction' procedure, has been shown to be very valuable in non-model species since it combines the advantages of cheap cost and high output (Rowe et al., 2011). RAD-Seq known to be used in about 20 species without using any reference genome in many phases of genomic/genetic research, comprising effective marker development approach (Pujolar et al., 2013), comparative mapping studies (Yang et al., 2013), QTL mapping (Hegarty et al., 2013), phylogenetic analyses (Nadeau et al., 2013) and genome-wide studies (Hecht et al., 2013). In species with existing reference genomes RAD-Seq has been a worthwhile substitute to WGRS (Whole Genome Re-Sequencing) for population genomic studies (Bruneaux et al., 2013; Varshney et al., 2013). Therefore, available RAD sequencing data provide new opportunities to develop SSR markers in bottle gourd and for several downstream analyses. Despite the significance of microsatellite markers in many research practices, genome-wide characterization of microsatellite sequences in the available bottle gourd genome has yet not been endeavored.

Till now only a few molecular markers, viz., 20 ISSR (Bhawna et al., 2014), 3226 SNP (Pei et al., 2013), SSRs (Xu et al., 2011; Bhawna et al., 2015a, b; Jacob et al., 2016; Mehtap et al., 2013; Wang et al., 2018), 54 RAPD (Decker-Walters et al., 2001), two chloroplasts and five nuclear markers (Andrew et al., 2006) have been available in bottle gourd to assess the crop genetic diversity. However, the existing resources are still insufficient for such a beneficial multifarious usage crop. Considering the efficacy of whole genome sequence-based microsatellite marker data, we aimed to develop microsatellite markers from RAD sequence data, and

supplement the distribution and density of perfect microsatellites in bottle gourd genome sequences. The work will prove to be useful to study phylogenetic, DNA fingerprinting and in breeding for bottle gourd and other cucurbits crop species, which have very less or no genomic markers.

2. Materials And Methods

2.1 Genomic sequences sourcing

The RAD assembled genome was obtained from the National Center for Biotechnology Information (NCBI), database search as *Lagenaria siceraria*, whole genome shotgun sequencing project or <http://www.ncbi.nlm.nih.gov/nucore/542029074>, or by using accession number ATBX01000001-ATBX01365668.

2.2 Detection of SSRs and primer design

The whole genome sequences of bottle gourd were searched for SSRs markers using MISA (MicroSatellite identification tool) tool available at <http://pgrc.ipk-gatersleben.de/misa/> with default parameters. The genomic flanking sequences of discovered microsatellite motifs were loaded into the Primer3 module of the MISA programme to design primer pairs with default parameters.

2.3 Validation of genomic SSRs

For validation of designed SSR markers, the genomic DNA was isolated from the leaves of 96 accessions of bottle gourd [Additional file 1, (Supplement Table S1)] from different region of India from the National Genebank of India, using the CTAB method. The isolated genomic DNA was separated on 0.8% agarose gel (Cambrex, USA) in Tris-borate-EDTA (TBE) buffer (pH 8.0) and quantified by means of Nanodrop, ND-1000 spectrophotometer (Wilmington, Delaware, USA) at 260 nm.

Genomic DNA amplification of the 96 accessions of bottle gourd germplasm was carried out using SSR primers in a total reaction volume of 12.3µl containing 30 ng/µl of DNA, 1X Taq polymerase reaction buffer, 2 mM MgCl₂, 0.1 mM each of dNTPs, 0.2 mM primer and 1 U of Taq DNA polymerase (MBI Fermentas, Germany) in a Bioer thermocycler (Hangzhou, P. R. China) with the following reaction conditions: initial denaturation for 5 min at 94°C followed by 35 cycles of 1 min denaturation at 94 °C, 1 min annealing

mostly in the range of 50-55°C, and 2 min extension at 72°C, followed by a final extension at 72°C for 7 min.

2.4 Genetic diversity and population structure of the National Genebank accessions of bottle gourd

After electrophoresis, the clear, intense and consistently amplified bands were scored for each accession in a binary matrix as 1 (presence) or 0 (absence) for diversity analysis study. The genetic relatedness was estimated by the genetic dissimilarity coefficient and the dendrogram were drawn using the unweighted pair group method (UPGMA) with NTSYS PC 2.1.

The binary data were also analyzed with POPGENE v. 1.32 (Yeh et al., 1999) to estimate different genetic diversity parameters, viz percentage of polymorphic loci (P), Shannon's information index (i), observed number of alleles (na), effective number of alleles (ne), and gene flow (Nm). GenAlex version 6.5 was used to assess the genetic diversity parameters. The polymorphic information content (PIC) and heterozygosity values were calculated for each SSR locus by using an online tool Gene-Calculator.

For the examination of population structure, we utilized a model-based clustering method as executed in the software program STRUCTURE 2.3.4 (Pritchard et al., 2000). To estimate the number of populations (K) and the admixture proportion for each germplasm, Monte Carlo Markov chain method was used. In this program, we used 1,00,000 iterations and a burn-in period of 1,00,000. At least ten independent runs were evaluated for each fixed number of populations (K).

2.5 Physical mapping of bottle gourd genomic microsatellite markers

For determining the physical location (bp) of identified microsatellite markers on the eleven bottle gourd chromosomes, the flanking genomic sequences of identified microsatellite repeat motifs were searched against the whole genome sequence of available bottle gourd, using BLASTn. Optimized BLASTn search parameters with default E-value (0.00001) and low-complexity filter selections were preferred for this analysis. Finally, the physically located microsatellite markers were visualized using MapChart software.

2.6 Comparative genome mapping between bottle gourd and other cucurbits species

The flanking sequences of microsatellite marker loci that were physically mapped on the eleven chromosomes of bottle gourd were BLAST searched against genome sequences of other *Cucurbitaceae* species like cucumber (*Cucumis sativus*) (Huang et al., 2009), melon (*Cucumis melo*) (Garcia-Mas et al., 2012) and

watermelon (*Citrullus lanatus*) (Guo et al., 2013) to derive marker-based syntenic relationships between bottle gourd and other *Cucurbitaceous* species. Sequences with 80% overlap and with no more than a 2-bp mismatch were only considered for analysis. A cutoff bit score of 54.7 and E-value, 1e205 were considered substantial for BLAST analysis. Syntenic relationships of bottle gourd based on SSR markers analyzed with cucumber, watermelon and melon, were finally visualized by means of software Circos 0.55 (Krzywinski et al., 2009) (<http://circos.ca>).

3. Results And Discussion

In the present study, we examined, available non-redundant RAD bottle gourd genomic sequence for the distribution of perfect microsatellites. Repeats were first examined on a whole genome basis using genomic DNA sequence of bottle gourd and microsatellite content in DNA sequences. Further, the SSR frequencies were studied and compared with cucumber and *Arabidopsis*. Mapping of genomic SSRs on linkage groups and, their comparative syntenic relationship was also studied with other *Cucurbitaceae* species.

3.1 Distribution of SSR types in genomic sequences

First, a total of 171,019 SSR markers with perfect repeats were detected in the bottle gourd; after removal of the redundant SSR markers, we finally found 45,066 perfect SSRs. The content of perfect microsatellites primers in genomic sequences of bottle gourd is summarized in Table 1. An overall density across the genome is 134.60 SSRs/Mbp (i.e., one SSR every 0.4 MB of sequence, excluding mononucleotide and compound SSRs). Expectedly, bottle gourd showed lowest microsatellite density among the species compared with that of cucumber (536.60 SSR/Mbp) and *Arabidopsis* (364.10 SSR/Mbp) (Table 2), this was due to lack of complete whole genome sequence data.

Tetranucleotide repeats were the most common SSR type in bottle gourd genomic sequence representing nearly 34% of all SSRs, followed by tri- (30.73%) and dinucleotides (21%) (Table 2). Penta and hexanucleotides were the least frequent repeat types, together representing less than 15% of the total SSRs. The distribution of SSR types in bottle gourd was most similar to that of cucumber, which had comparable relative and absolute frequencies for almost each SSR type, as compared to *Arabidopsis*. In bottle gourd tetranucleotides

were predominant as in cucumber whereas *Arabidopsis* has trinucleotides as the abundant SSR types. The distribution of bottle gourd microsatellites with respect to the number of repeat units is shown in Fig. 1. For all SSR types, microsatellite frequency declined with increase in the number of repeat units. However, the frequency of this variation was more gradual in all nucleotides from di- to tetranucleotides than in longer repeat types, from penta- to hexanucleotides showing the most dramatic reduction in frequency with increased repeat units (Table 2). For instance, the mean number of repeat units in dinucleotides are 9.0, which were twice as that of the number of repeat units in tetranucleotides (4.5) (Table 2); dinucleotides also revealed the presence of more varied motifs despite less number of SSRs than tetranucleotides. Although tetra, penta and hexa having almost the same mean repeat number like 4.5, 4 and 4 respectively, showing equal contribution to the genome portion that is occupied by SSRs nucleotides. Trinucleotides repeat motif occurred more frequently than tetra, penta and hexanucleotide in the RAD genome data with a mean of 6.5.

3.2 Distribution of microsatellite motifs

We have carried out a detailed analysis of individual repeat motifs for each type of SSRs found in genomic sequences of bottle gourd, along with related calculations used for cucumber and *Arabidopsis*, which is presented as supplemental data [Additional file 1, Supplement Table S2 (for bottle gourd), Supplement Table S3 (for both cucumber and *Arabidopsis*)] available with the online version of this paper. The relative frequency (%) of SSR types with different number of repeats, present in the bottle gourd genome represent in Fig. 2. The main results of this analysis were outlined for different motifs separately.

Dinucleotide motifs: In bottle gourd, genomic sequences, analysis of the di nucleotide sequence shows AT motif was predominantly present (Additional file 1, Table S2) and CG repeats were the least frequent dinucleotides in bottle gourd. In addition, relate to other both species, the existence of AT also being the amplest single motif, regardless of repeat type. Whereas, GC showed the least frequent repeat in both cucumber and *Arabidopsis*, in all across the genomic data sets examined. Comprehensive frequencies of individual dinucleotide motifs are mentioned in Additional file 1, Table S3.

Trinucleotide motifs: In bottle gourd, trinucleotide repeats were the second most abundant in genomic

sequences available. Investigating frequencies of diverse trinucleotides revealed that repeats of AAT were more common in bottle gourd genomic data along with cucumber. But this pattern is not like that of *Arabidopsis*, as in *Arabidopsis* AAG is predominant one. Conversely, CGT were the rarest trinucleotides in genomic DNA of bottle gourd whereas CCG shows the least abundant in both cucumber and *Arabidopsis*.

Tetranucleotide motifs: Tetranucleotide repeats were the most frequent in genomic sequences of bottle gourd (Table 2). The AT-rich motifs AAAT, TTTA, ATTT, TTAT and AATA were, the most abundant tetranucleotides in bottle gourd genomic data, together representing ~62% of all tetramer repeats. Whereas GC-rich repeats like ACGC, AGGC, CCCG, GCAG, GCCT, GCGG etc. were the least, with relative frequencies of >0.1% (Additional file 1, Table S2). A comparable distribution was perceived in the other dicots also, showing a clear prevalence of these similar AT-rich motifs in study.

Pentanucleotide motifs: Bottle gourd genomic sequences had the least frequency of pentanucleotide repeats. Among them AT rich penta repeats were the predominant one accounting a total of ~51% penta repeats most common motifs among them are AAAAT, ATTTT, TTTTA, TAAAA, TTTAT etc. followed by AAAAG (see Additional file 1, Table S2). In general, these motifs also predominated in both other species, akin to cucumber with the highest density of AAAAG followed by AAAAT repeats in genomic sequences, whereas *Arabidopsis* had the highest frequency of AAAAT motif followed by AAAAC, AAAAG. Analysis of pentanucleotide frequencies in genomic DNA from all three-species revealed that, in both bottle gourd and *Arabidopsis*, AAAAT was the most abundant repeat, with the next most frequent repeats, AAAAG or AAAAC, but with cucumber AAAAG is more dominant motif than the AAAAT.

Hexanucleotide motifs: In bottle gourd, genomic sequence least frequent SSRs are the hexanucleotide accounting only 4.3% among all SSRs. AT-rich hexanucleotide motifs such as AAAAAT, ATTTTT, TATTTT, AAAATA, TTATTT etc. together contribute to 43.38% in bottle gourd genomic data. In all motifs AAAAAT is the most frequent repeat present followed by AAAAAG, TAAAAA, contributing 43.38% of total hexanucleotide. Similarly, *Arabidopsis* also has AAAAAT as the predominant one and in cucumber AAAAAG followed by AAAAAT are frequently present.

3.3 Development of microsatellite markers and their physical mapping in bottle gourd genome

Initially we identified 1,70,775 SSRs markers, from the RAD genomic sequence of bottle gourd. After removing redundant and compound SSRs, we finally developed 45,066 perfect SSRs. These SSRs had forward and reverse primers on the both sides of the SSR flanking genomic sequences (Supplementary Table S4) and had 100% successful primer designing potential. All scaffolds were not mapped on chromosomes and only 922 scaffolds/contigs were effectively anchored to the genetic map (Pei et al., 2013), therefore only 207 mapped microsatellites could be developed. The genomic distribution of these 207 microsatellite markers in the bottle gourd genome revealed their physical localization on the 11 linkage groups of bottle gourd with average marker density of 0.62 markers per MB (Fig. 3). The average marker density was maximum 0.29/Mb in chromosome 5, followed by 0.25/Mb in chromosome 7. From these data, no SSRs were anchored over linkage group 3 and 11 but linkage group 10, 4, 8 and 1 shows 0.03/Mb, 0.02/Mb, 0.01/Mb and 0.01/Mb average marker density respectively, and with least number of markers in chromosome 2, 6 and 9 (0.002/Mb). These observations suggested that this data is not high-density microsatellite marker mapping, but the data was efficiently utilized to give maximum knowledge based on the available RAD genomic sequence. Herein, all genome-wide SSR markers, including both physically mapped and non-mapped markers for bottle gourd genome were developed. This huge marker data could be advantageous for various applications, like large-scale genotyping, phylogenetic relationship studies, molecular breeding, diversity analysis, population genetic structure analysis, including mapping of genes and comparative genome mapping involving bottle gourd and other cucurbit crop plants.

3.4 Amplification and polymorphic potential of microsatellite markers

The exact positions of SSRs in the bottle gourd RAD sequenced genome, their primers as well as information on expected PCR product length, repeat motifs and primers AT (annealing temperature), are presented in Additional files 1, Supplementary Table S4. 203 primer pairs of random SSRs motif length (Additional files 1, Supplementary Table S5) were validated using PCR and 49.75% polymorphic primer pairs were seen and remaining loci were monomorphic.

3.5 SSR polymorphism and Genetic diversity analysis

The alleles recognized from the 120 novel polymorphic marker pairs varied from 2 to 8 with a mean of 2.43 alleles per locus. The PIC values for these polymorphic SSR markers varied from 0.02 (Lasi 4) to 0.83 (Lasi 42) with a mean of 0.45 (Table 4). UPGMA dendrogram was used to analyse genetic diversity analysis by using polymorphic markers, based on the distinctive DNA marker profiles (Fig. 4) of all the National Genebank accessions of bottle gourd used in the study. Scored polymorphic data were subjected to NTSYS- PC 2.01 software for generating the UPGMA dendrogram based on the Jaccard's similarity coefficient and the 96 accessions were distributed in two clusters. One cluster was of small size with seven germplasm from Delhi region and another large cluster contained 5 sub-clusters and a few out-groups (Fig. 5). In sub-cluster 1, maximum germplasm was from Delhi region except two IC-317498 from Bihar and IC-341128 from Uttarakhand. Whereas sub-cluster 2, 3, 4 and 5 having germplasm from different part of India. The outgroup IC-275 92390, IC-430102, IC-0536593, IC-522210, IC-519467, IC-301188, IC-92807, IC-0385816 and IC-0382208 representing germplasm separated from the other clusters, these diverse accessions may be due to the presence of some unique traits and could be utilized in breeding programs. The obtained pairwise genetic similarity matrix of Jaccard's coefficient calculate, genetic similarity (GS) ranged from 0.77 to 0.99, with a mean of 0.87. The minimum similarity value, 0.77, between IC-92376 and IC-52220 and the maximum similarity value 0.99 was scored between IC-33833 and IC-3428, suggested the germplasm is not much diverged.

Various diversity parameters like gene diversity, Shannon's information index, gene flow, number of alleles etc. with each of these 120 polymorphic markers are shown in Table 5. Observed number of alleles and effective number of alleles were 1.99 and 1.45 respectively among the studied accessions of bottle gourd. Nei's Gene Diversity which is equivalent to the average heterozygosity was 0.25 ± 0.21 and Shannon's Information Index was 0.38 ± 0.27 . These results indicated that the accessions harbour a less diversity, High gene flow value was obtained, i.e., 0.25 for the populations used. This gene flow value is possibly because of the cross-pollinated nature of the crop and that no boundaries on genetic exchange within India.

3.6 Population structure analyses

A model-based clustering analysis was performed for all 96 accessions using 120 polymorphic SSR

markers. Ad hoc quantity (ΔK) was used to solve the difficulty in inferring the real K value. A comparatively high value of ΔK was found for $K = 4$ (Fig. 6). By analyzing the 96 bottle gourd accessions using the STRUCTURE program (Fig. 7), two germplasm IC-92469 and IC-505650 from Delhi clustered with other germplasm of sub-cluster 1 (Blue) and also showed admixing with another sub-cluster of Delhi region (Green). Subcluster-2 and 3 showed varying degree of admixture population. The genetic structuring events between germplasm within populations is a result of the interaction of the various evolutionary forces, mutation, migration, breeding structure, and selection pressure that may influence simultaneously or discreetly. In the UPGMA clustering subcluster-4 and 5 are separated from each other whereas STRUCTURE merge them into one population.

3.7 Comparative genome mapping between bottle gourd and other cucurbits species

For comparative mapping, the physically mapped 207 microsatellite markers of bottle gourd were associated with their physical position on the chromosomes of other correlated cucurbit genomes namely cucumber, melon and watermelon (Fig. 8). Comparative genome mapping shown a maximum fraction of sequence established orthology. Therefore, microsatellite markers scattered over 11 bottle gourd linkage groups revealed correlation with 7 chromosomes of cucumber having 93.2% synteny, 12 chromosomes of melon with 87.4% and 11 chromosomes of watermelon with 90.8% synteny was identified.

Bottle gourd–cucumber synteny: 193 bottle gourd SSRs were mapped between cucumber and bottle gourd and the markers in all 11 bottle gourd linkage groups revealed an average frequency of 14.28 % to that of marker chromosomal location in cucumber. The physically mapped microsatellite markers on the bottle gourd chromosome 5 (90) showed maximum synteny of 21.1% with cucumber chromosome 5, followed by between bottle gourd chromosome 5 and cucumber chromosome 4 i.e., 18.8%. Bottle gourd chromosome 7 having synteny of 77 microsatellites with cucumber chromosome 4, 5 and 6 which having 20.7%, 18.2% and 24.6% synteny respectively.

Bottle gourd–watermelon synteny: The comparative mapping between bottle gourd and watermelon genomes shown a syntenic relationship of 188 microsatellite marker loci dispersed over 11 linkage groups of bottle gourd and 11 chromosomes of watermelon. It was identified that a synteny relationship of 12.2% genomic

SSRs on 10 and 11 chromosomes of watermelon each, and followed by 11.7% genomic SSRs on watermelon chromosome number 8 and the minimum between bottle gourd chromosome 2, 6 and 9 and watermelon chromosome 4. On an average, 9.09% syntenic relationship of SSR marker loci between bottle gourd and watermelon chromosomes was identified.

Bottle gourd–melon synteny: The SSR markers physically mapped on the bottle gourd linkage groups showed least synteny with 3 and 10 melon chromosomes with an average frequency of 8.3% that is much lower than that with the both cucumber and watermelon chromosomes. About 181 bottle gourd microsatellite marker loci showed significant matches with genomic regions spanning over 12 chromosomes of melon genome. Highest synteny of SSRs marker loci between bottle gourd chromosome 5 and melon chromosome 5 (89.5%) and minimum synteny observed in bottle gourd chromosome 2, 6, 9 and melon chromosome 3 and 10. We are reporting huge numbers of SSR markers in bottle gourd utilizing RAD sequence data. The current study also provided some interesting information like syntenic relationships of SSR markers between bottle gourd, cucumber, melon and watermelon. Using synteny study, an idea of genome conservation among cucurbits like cucumber, melon and watermelon was revealed and the comparative studies are valuable in understanding the evolutionary progression among cucurbits genome..

4. Conclusion

RAD sequencing information was proved to be highly valued in large-scale development of SSRs markers. In our study, this valuable data had added to a detailed representation of microsatellites in the bottle gourd genome. Here, we revealed the genome wide identification of 45,066 microsatellite motifs in the bottle gourd, where tetranucleotide repeats were the utmost frequent SSRs present in the genome. Due to limited availability of assembled RAD genome sequence, only 207 SSRs were physically mapped on 11 linkage groups of bottle gourd genome and were further used to develop synteny with cucumber, melon and watermelon. Among 45,066 microsatellites primer pairs of SSRs, all mapped markers were synthesized and successfully validated with quality amplification and polymorphic potential in bottle gourd accessions. These markers were then utilized for genetic diversity evaluation of the bottle gourd Indian National Genebank germplasm. These

reported SSR markers would be of enormous use for germplasm categorization, for gene or QTL detection, cultivar identification and relative genome mapping involving bottle gourd and other cucurbits species. The microsatellite marker based comparative genome mapping amongst bottle gourd and other cucurbits species such as cucumber, melon and watermelon would be advantageous in a map based isolation of agronomic importance genes from the bottle gourd by means of the marker-based genotyping data.

Declarations

Acknowledgements

Grateful to Pei Xu, China for timely assistance in data acquisition. The first author BB acknowledges the award of Senior Research Fellowship from University Grants Commission, New Delhi. Support from National Bureau of Plant Genetic Resources, New Delhi, India is duly acknowledged.

Declaration of Interest Statement

The authors declare no conflicts of interest for this article.

References

- Altshuler, D., Pollara, V.J., Cowles, C.R., Van, E.W.J., Baldwin, J., Linton, L., Lander, E.S., 2000. An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature*. 407, 513–516.
- Andrew, C.C., Michael, K.B., Patricia, A.M., David, L.E., David, P., 2006. Reconstructing the Origins and Dispersal of the Polynesian Bottle Gourd (*Lagenaria siceraria*). *Molecular Biology and Evolution*. 23,893–900.
- Badmanaban, R., Patel, C.N., 2010. Studies on anthelmintic and antimicrobial activity of the leaf extracts of *Lagenaria siceraria*. *Molecular Journal of Global Pharma Technology*. 4, 66–70.
- Barot, A.M., Suneeta, P., Hiral, M., 2014. Development of Technology for Manufacture of Bottle Gourd IceCream. *J Nutr Food Sci*. 4, 6.
- Bhawna, Abdin, M.Z., Arya, L., Dipnarayan, S., Sureja, A.K., Chitra, P., Verma, M., 2014. Population structure and genetic diversity in bottle gourd [*Lagenaria siceraria* (Mol.) Standl.] germplasm from India assessed by ISSR markers. *Plant Systematics and Evolution*. 300, 767-773.
- Bhawna, Abdin, M.Z., Arya, L., Verma, M., 2015. Development of novel gene-based microsatellite markers for robust genotyping purposes in *Lagenaria siceraria*. *Scientia Horticulturae*. 191, 15–24.

- Bhawna, Abdin, M.Z., Arya, L., Verma, M., 2015. Transferability of cucumber microsatellite markers used for phylogenetic analysis and population structure study in bottle gourd (*Lagenaria siceraria* (Mol.) Standl.). *Appl. Biochem. Biotechnol.* 175, 2206–2223.
- Bowers, J.E., Meredith, C.P., 1997. The parentage of a classic wine grape, Cabernet Sauvignon. *Nature Genetics.* 16, 84-87.
- Bruneaux, M., Johnston, S.E., Herczeg, G., Merila, J., Primmer, C.R., Vasemagi, A., 2013. Molecular evolutionary and population genomic analysis of the nine-spined stickleback using a modified restriction-site-associated DNA tag approach. *Molecular Ecology.* 22, 5065–5082.
- Danieli, M.N., Lauren, d.C.D., 2020. Waterproofing of bottle gourd (*Lagenaria siceraria*) with castor oil polyurethane resin. *Belo Horizonte.* 1 – 4.
- Decker-Walters, D., Staub, J., Lopez-Sese, A., Nakata, E. 2001. Diversity in landraces and cultivars of bottle gourd (*Lagenaria siceraria*; Cucurbitaceae) as assessed by random amplified polymorphic DNA. *Genetic Resources and Crop Evolution.* 48, 369–380.
- Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E., 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE.* 6,e19379.
- Emina, M., Berenji, J., Ognjanov, V., Mirjana L., Jelena, C., 2012. Genetic variability of bottle gourd *Lagenaria siceraria* (mol.) standley and its morphological characterization by multivariate analysis. *Arch. Biol. Sci., Belgrade,* 64(2), 573-583.
- Erickson, D.L., Smith, B.D., Clarke, A.C., Sandweiss, D.H., Tuross, N., 2005. An Asian origin for a 10,000-year-old domesticated plant in the Americas. *Proceedings of the National Academy of Sciences USA.* 102, 18315–18320.
- Fard, M.H., Bodhankar, S.L., Dikshit, M., 2008. Cardioprotective activity of fruit of *Lagenaria siceraria* (Molina) Standley on Doxorubicin induced cardiotoxicity in rats. *International Journal of Pharmacology.* 6, 466–471.
- Garcia-Mas, J., Benjak, A., Sanseverino, W., 2012. The genome of melon (*Cucumis melo* L.). *Proceedings of the National Academy of Sciences USA.* 109, 11872–11877.
- Garza, J.C., Slatkin, M., Freimer, N.B., 1995. Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. *Molecular Biology and Evolution.* 12, 594-603.

- Guo, S., Zhang, J., Sun, H., 2013. The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nature Genetics*. 45, 51–58.
- Hecht, B.C., Campbell, N.R., Holecek, D.E., Narum, S.R., 2013. Genome-wide association reveals genetic basis for the propensity to migrate in wild populations of rainbow and steelhead trout. *Molecular Ecology*. 22, 3061–3076.
- Hegarty, M., Yadav, R., Lee, M., Armstead, I., Sanderson, R., Scollan, N., Powell, W., Skøt, L., 2013. Genotyping by RAD sequencing enables mapping of fatty acid composition traits in perennial ryegrass [*Lolium perenne* (L.)]. *Plant biotechnology*. 11, 572–581.
- Heiser, C.B., 1979. The gourd book. University of Oklahoma Press, Norman, Oklahoma. A thorough and fascinating account of gourds from throughout the world.
- Helyar, S.J., Hemmer-Hansen, J., Bekkevold, D., et al. 2011. Application of SNPs for population genetics of non-model organisms: new opportunities and challenges. *Molecular Ecology Resources*. 1, 123–136.
- Huang, S.W., Li, R.Q., Zhang, Z.H., et al., 2009. The genome of the cucumber, *Cucumis sativus* L. *Nature Genetics*. 41, 1275–1281.
- Innan, H., Terauchi, R., Miyashita, N.T., 1997. Microsatellite polymorphism in natural populations of the wild plant *Arabidopsis thaliana*. *Genetics*. 146, 1441-1452.
- Jacob, M., Hussein, A.S., Alfred, O.O., Beyene, A.A., 2016. Amelework Genetic differentiation of bottle gourd [*Lagenaria siceraria* (Molina) Standl.] landraces assessed by fruit qualitative traits and simple sequence repeat markers. *Scientia Horticulturae*. 216, 1-11.
- Kashi, Y., King, D., Soller, M., 1997. Simple sequence repeats as a source of quantitative genetic variation. *Trends in Genetics*. 13, 74-78.
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., Marra, M.A. 2009. Circos: an information aesthetic for comparative genomics. *Genome Research*, 19, 1639–1645.
- Li, Y.C., Korol, A.B., Fahima, T., Beiles, A., Nevo, E., 2002. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology*. 11, 2453-2465.
- MacHugh, D.E., Shriver, M.D., Loftus, R.T., Cunningham, P., Bradley, D.G., 1997. Microsatellite DNA

variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics*. 146, 1071–1086.

McCouch, S.R., Teytelman, L., Xu, Y.B., Lobos, K.B., Clare, K., Walton, M., Fu, B.Y., Maghirang, R., Li, Z.K., Xing, Y.Z., Zhang, Q.F., Kono, I., Yano, M., Fjellstrom, R., DeClerck, G., Schneider, D.,

Cartinhour, S., Ware, D., Stein, L., 2002. Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). *DNA Research*. 9, 199-207.

Mehtap, Y., Hugo, E.C., Suat, S., Ceknas, E., Faheem, S.B., 2013. Transferability of Cucurbita SSR markers for genetic diversity assessment of Turkish bottle gourd (*Lagenaria siceraria*) genetic resources.

Biochemical Systematics and Ecology. 59, 45-53.

Morgante, M., Hanafey, M., Powell, W., 2002. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genetics*. 30, 194-200.

Nadeau, N.J., Martin, S.H., Kozak, K.M., Salazar, C., Dasmahapatra, K.K., Davey, J.W., Baxter, S.W.,

Blaxter, M.L., Mallet, J., Jiggins, C.D., 2013. Genome-wide patterns of divergence and gene flow across a butterfly radiation. *Molecular Ecology*. 22, 814–826.

Nidhi, T., Ganesh, N.S., Bhirendra S., 2017. Medicinal value of *Lagenaria siceraria*: An overview. *Int. J. Indig. Herb Drug*. 36-43.

Palamthodi, S., Kadam, D., Lele, S.S., 2019. Physicochemical and functional properties of ash gourd/bottle gourd beverages blended with jamun. *J. Food Sci. Technol*. 56(1), 473–482.

Pei, X., Shizhong, X., Xiaohua, W., Ye, T., Baogen, W., Sha, W., Dehui, Q., Zhongfu, L., Guojing, Li., 2013.

Population genomic analyses from low-coverage RAD-Seq data: a case study on the non-model cucurbit bottle gourd. *The Plant Journal*. 77, 430-442.

Powell, W., Morgante, M., Andre, C., Henfey, M., Vogel, J., Tingy, S., Rafalsky, A., 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding*. 2, 225-238.

Prashant, P.S., Rama, C.P., Birendra, N.R., 2014. A green separation of *Lagenaria siceraria* seed oil. *Industrial Crops and Products*. 52, 796–800.

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

Pujolar, J.M., Jacobsen, M.W., Frydenberg, J., Als, T.D., Larsen, P.F., Maes, G.E., Zane, L., Jian, J.B., Cheng, L., Hansen, M.M., 2013. A resource of genome-wide single-nucleotide polymorphisms generated by RAD tag sequencing in the critically endangered European eel. *Molecular Ecology Resources*. 13, 706–714.

Rowe, H.C., Renaut, S., Guggisberg, A., 2011. RAD in the realm of next-generation sequencing technologies. *Molecular Ecology*. 20, 3499–34502.

Sakshi, M., Akhilesh, T., Sanyam, G., Ajay, S., Anil, K.G., 2015. An overview on *Lagenaria siceraria* (bottle gourd). *Journal of Biomedical and Pharmaceutical Research*. 4(3), 04-10.

Santi, L., Wang, Y., Stile, M.R., Berendzen, K., Wanke, D., Roig, C., Pozzi, C., Muller, K., Muller, J., Rohde, W., Salamini, F., 2003. The GA octodinucleotide repeat binding factor BBR participates in the transcriptional regulation of the homeobox gene *Bkn3*. *Plant*. 34, 813-826.

Saveliev, A., Everett, C., Sharpe, T., Webster, Z., Festenstein, R., 2003. DNA triplet repeats mediates heterochromatin-protein-1-sensitive variegated gene silencing. *Nature*. 422, 909-913.

Somers, D.J., Isaac, P., Edwards, K., 2004. A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.) *Theoretical and Applied. Genetics*. 109, 1105-1114.

Subramanian, S., Mishra, R.K., Singh, L., 2003. Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. *Genome Biology*. 4, R13.

Tautz, D., Renz, M., 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research*. 12, 4127–38.

Toth, G., Gaspari, Z., Jurka, J., 2000. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Research*. 10, 967-981.

Upaganlawar, A., Balaraman, R., 2010. Protective effects of *Lagenaria siceraria* (Molina) fruit juice in isoproterenol induced myocardial infarction. *International Journal of Pharmacology*. 5, 645–651.

Varshney, R.K., Graner, A., Sorrells, M.E. 2005. Genic microsatellite markers in plants: features and

applications. *Trend in Biotechnolgy* 23, 48-55.

Varshney, R.K., Song, C., Saxena, R.K., 2013. Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nature Biotechnology*. 31, 240–246.

Wang, Y., Xu, P., Wu, X., Wu, X., Wang, B., Huang, Y., Hu, Y., Lin, J., Lu, Z., Li, G., 2018. GourdBase: A genome-centered multi-omics database for the bottle gourd (*Lagenaria siceraria*), an economically important cucurbit crop. *Scientific Reports*. 8, 3604.

Whitaker, T.W., 1971. Endemism and pre-Columbian migration of bottle gourd, *Lagenaria siceraria* (Mol.) Standl. *Man across the sea*. 78-218.

Xu, P., Wu, X., Luo, J., Wang, B., Liu, Y., Ehlers, J.D., Wang, S., Lu, Z., Li, G., 2011. Partial sequencing of the bottle gourd genome reveals markers useful for phylogenetic analysis and breeding. *BMC Genomics*. 120, 467.

Yang, H., Tao, Y., Zheng, Z., Zhang, Q., Zhou, G., Sweetingham, M.W., Ho-wieson, J.G., Li, C., 2013. Draft genome sequence, and a sequence defined genetic linkage map of the legume crop species *Lupinus angustifolius* L. *PLoS ONE*. 8, e64799.

Yeh, F.C., Yang, R.C., Boyle, T., 1999. POPGENE Version 1.31. Microsoft Window-based Freeware for Population Genetic Analysis.

Tables

Table 1 Sum total of different distributed perfect microsatellites primers with different motif repeats length in genomic sequences of bottle gourd.

	Motif Length	Primer Numbers
1	12	27253
2	14	2533
3	15	6942
4	16	3463
5	18	3402
6	20	908
7	21	239
8	22*	74
9	24	221
10	25*	23
11	27*	1
12	30*	7

(* indicates synthesized primer used to test in bottle gourd accessions.)

Table 2 Distribution of perfect microsatellites with their relative frequency, mean repeat number in genomic sequences of bottle gourd and microsatellite density comparison with *Arabidopsis thaliana* and *Cucumis sativus*.

Sequence type	Count	Rel. freq. (%)	Mean repeat number	Microsatellite density (SSR/Mbp)		
<i>Genomic type</i>	<i>SSRs</i>	<i>Lagenaria</i>	<i>siceraria</i>	<i>Lagenaria</i> <i>Cucumis siceraria</i>	<i>Arabidopsis</i> <i>sativus</i>	<i>thaliana</i>
Dinucleotide	9463	21.00	9.00	28.30	78.70	146.00
Trinucleotide	13850	30.73	6.50	41.40	146.60	141.00
Tetranucleotide	15480	34.35	4.50	46.30	93.20	164.20
Pentanucleotide	4337	9.62	4.00	12.90	32.00	54.40
Hexanucleotide	1936	4.30	4.00	5.70	13.60	31.00
Total				134.60	364.10	536.60
Total/mean	45066	100.00	5.60	26.92	72.82	107.32
GC content (%)				37.50	36.00	41.40

Table 3 SSR motifs identified in different type of SSRs repeat nucleotides.

	SSR Motif Type	Dinucleotide	Trinucleotide	Tetranucleotide	Pentanucleotide	Hexanucleotide
No of Total Repeats In Different Motifs	2	0	0	0	0	0
	3	0	0	13076	3907	1768
	4	0	9511	2063	407	161
	5	0	3035	317	23	7
	6	4666	1038	24	0	0
	7	2533	239	0	0	0
	8	1400	26	0	0	0
	9	596	1	0	0	0
	10	184	0	0	0	0
	11	74	0	0	0	0
	12	10	0	0	0	0
	Total	9463	13850	15480	4337	1936
	%	21.00	30.73	34.35	9.62	4.30

Table 4 List of 120 SSR polymorphic markers screened with bottle germplasm with various genetic diversity parameters like number of alleles, gene diversity, Shannon's information index, gene flow, etc. in 96 bottle gourd accessions.

Primer Name	Sample Size	ns*	1*	Obs_Hom	Obs_Het	Exp_Hom*	Exp_Het*	Nu**	Ave_Het
Bg5SR1050	192	2	1.04	0.1	0.96	0.04	0.96	0.04	0.04
Bg5SR1135	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR1485	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR1484	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR2066	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR2267	188	2	1.99	0.69	0.05	0.95	0.5	0.5	0.5
Bg5SR2302	192	2	1.11	0.2	1	0	0.9	0.1	0.1
Bg5SR2784	180	2	1.07	0.15	1	0	0.94	0.06	0.06
Bg5SR3245	192	2	1.77	0.61	0.64	0.36	0.56	0.44	0.43
Bg5SR3318	192	2	1.05	0.12	0.95	0.05	0.95	0.05	0.05
Bg5SR4720	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR4910	190	2	1.92	0.67	0.2	0.8	0.52	0.48	0.48
Bg5SR5558	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR5629	192	2	1.27	0.37	0.76	0.24	0.79	0.21	0.21
Bg5SR5652	192	2	1.65	0.58	1	0	0.6	0.4	0.4
Bg5SR5677	192	2	1.08	0.16	0.93	0.07	0.93	0.07	0.07
Bg5SR7093	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR8718	192	2	1.31	0.4	0.73	0.27	0.76	0.24	0.23
Bg5SR9791	192	2	1.19	0.3	0.82	0.18	0.84	0.16	0.16
Bg5SR9588	190	2	1.99	0.69	0.07	0.93	0.5	0.5	0.5
Bg5SR10214	192	1	1	0	1	0	1	0	0
Bg5SR10512	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR10616	192	2	1.03	0.08	0.97	0.03	0.97	0.03	0.03
Bg5SR11085	192	2	1.03	0.08	0.97	0.03	0.97	0.03	0.03
Bg5SR11162	192	2	1.09	0.17	0.92	0.08	0.92	0.08	0.08
Bg5SR11414	192	2	1.02	0.08	0.98	0.02	0.98	0.02	0.02
Bg5SR11815	188	2	1.03	0.08	0.97	0.03	0.97	0.03	0.03
Bg5SR12456	192	2	1.1	0.19	0.91	0.09	0.91	0.09	0.09
Bg5SR13004	184	2	1.91	0.67	0.22	0.78	0.52	0.48	0.48
Bg5SR13008	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR13442	192	2	1.44	0.48	0.98	0.02	0.98	0.31	0.3
Bg5SR14143	192	2	1.02	0.06	0.98	0.02	0.98	0.02	0.02
Bg5SR15149	186	2	1.99	0.69	0.09	0.91	0.5	0.5	0.5
Bg5SR15165	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR16207	192	2	1.96	0.68	0.14	0.86	0.51	0.49	0.49
Bg5SR16402	192	2	1.56	0.43	1	0	0.73	0.27	0.26
Bg5SR16406	186	2	1.17	0.28	0.97	0.03	0.95	0.15	0.15
Bg5SR17040	192	2	1.01	0.03	0.99	0.01	0.99	0.01	0.01
Bg5SR17091	192	2	1.06	0.14	0.96	0.04	0.94	0.06	0.06
Bg5SR17146	192	2	1.04	0.1	0.96	0.04	0.96	0.04	0.04
Bg5SR17160	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR17238	192	2	1.08	0.16	0.93	0.07	0.93	0.07	0.07
Bg5SR17321	192	2	1.25	0.36	0.77	0.23	0.8	0.2	0.2
Bg5SR17421	192	2	1.96	0.68	0.15	0.85	0.51	0.49	0.49
Bg5SR17443	192	2	1.96	0.68	0.14	0.86	0.51	0.49	0.49
Bg5SR17624	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR18211	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR18353	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR19293	192	2	1.01	0.03	0.99	0.01	0.99	0.01	0.01
Bg5SR19686	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR20604	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR21066	192	2	1.03	0.08	0.97	0.03	0.97	0.03	0.03
Bg5SR21687	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR21730	190	2	1.99	0.69	0.48	0.54	0.5	0.5	0.5
Bg5SR21884	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR22073	192	2	1.74	0.62	0.76	0.24	0.57	0.43	0.43
Bg5SR22928	190	2	1.01	0.03	0.99	0.01	0.99	0.01	0.01
Bg5SR23156	192	2	1.49	0.51	0.58	0.42	0.67	0.33	0.33
Bg5SR24331	192	2	1.02	0.06	0.98	0.02	0.98	0.02	0.02
Bg5SR24393	192	2	1.04	0.1	0.98	0.02	0.96	0.04	0.04
Bg5SR24889	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR24937	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR27153	192	2	1.09	0.17	1	0	0.92	0.08	0.08
Bg5SR28137	192	2	1.01	0.03	0.99	0.01	0.99	0.01	0.01
Bg5SR28162	192	2	1.05	0.12	0.95	0.05	0.95	0.05	0.05
Bg5SR28906	192	2	1.23	0.33	0.98	0.02	0.81	0.19	0.19
Bg5SR28922	192	2	1.99	0.69	0.06	0.94	0.5	0.5	0.5
Bg5SR28992	192	2	2	0.69	0.01	0.99	0.5	0.5	0.5
Bg5SR29684	192	2	1.41	0.47	1	0	0.71	0.29	0.29
Bg5SR29801	190	2	1.71	0.61	0.43	0.57	0.58	0.42	0.42
Bg5SR29900	192	2	1.05	0.12	0.95	0.05	0.95	0.05	0.05
Bg5SR30770	192	2	1.01	0.03	0.99	0.01	0.99	0.01	0.01
Bg5SR30515	192	2	1.02	0.06	0.98	0.02	0.98	0.02	0.02
Bg5SR30724	190	2	1.97	0.69	1	0	0.5	0.5	0.49
Bg5SR30956	192	2	1.08	0.16	0.93	0.07	0.93	0.07	0.07
Bg5SR31611	190	2	1.8	0.64	0.99	0.01	0.55	0.45	0.44
Bg5SR31828	192	2	1.09	0.17	0.92	0.08	0.92	0.08	0.08
Bg5SR32062	192	2	1.75	0.62	0.69	0.31	0.57	0.43	0.43
Bg5SR32118	190	2	1.59	0.56	0.72	0.28	0.63	0.37	0.37
Bg5SR32484	190	2	1.34	0.42	0.77	0.23	0.75	0.25	0.25
Bg5SR32529	192	2	1.01	0.03	0.99	0.01	0.99	0.01	0.01
Bg5SR32652	192	2	1.12	0.22	0.89	0.11	0.89	0.11	0.11
Bg5SR32882	192	2	1.05	0.12	0.95	0.05	0.95	0.05	0.05
Bg5SR33063	192	2	1.05	0.12	0.95	0.05	0.95	0.05	0.05
Bg5SR33219	192	2	1.49	0.51	0.58	0.42	0.67	0.33	0.33
Bg5SR33437	192	2	1.48	0.5	0.67	0.01	0.67	0.33	0.32
Bg5SR34029	192	2	1.01	0.03	0.99	0.01	0.99	0.01	0.01
Bg5SR34203	192	2	1.01	0.03	0.99	0.01	0.99	0.01	0.01
Bg5SR34310	192	2	1.02	0.06	1	0	0.98	0.02	0.02
Bg5SR34447	192	2	1.01	0.03	0.99	0.01	0.99	0.01	0.01
Bg5SR34657	192	2	1.02	0.06	0.98	0.02	0.98	0.02	0.02
Bg5SR35166	192	2	1.03	0.08	0.97	0.03	0.97	0.03	0.03
Bg5SR35748	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR35808	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR35816	192	2	1.01	0.03	0.99	0.01	0.99	0.01	0.01
Bg5SR35826	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR36440	188	2	1.45	0.49	0.62	0.38	0.69	0.31	0.31
Bg5SR36547	192	2	1.93	0.68	0.19	0.81	0.52	0.48	0.48
Bg5SR36577	192	2	1.01	0.03	0.99	0.01	0.99	0.01	0.01
Bg5SR36586	190	2	1.7	0.6	0.61	0.39	0.59	0.41	0.41
Bg5SR37184	192	2	1.05	0.12	0.97	0.03	0.95	0.05	0.05
Bg5SR37191	192	2	1.05	0.12	0.95	0.05	0.95	0.05	0.05
Bg5SR37324	184	2	1.97	0.68	0.96	0.04	0.51	0.49	0.49
Bg5SR37566	192	2	1.19	0.3	0.84	0.16	0.84	0.16	0.16
Bg5SR37687	190	2	1.02	0.06	0.98	0.02	0.98	0.02	0.02
Bg5SR37835	192	2	1.17	0.27	0.99	0.01	0.86	0.14	0.14
Bg5SR38173	189	2	1.41	0.57	0.51	0.49	0.62	0.38	0.38
Bg5SR38439	192	2	1.02	0.06	0.98	0.02	0.98	0.02	0.02
Bg5SR38559	188	2	1.02	0.06	1	0	0.98	0.02	0.02
Bg5SR41800	192	2	1.02	0.06	1	0	0.98	0.02	0.02
Bg5SR42187	192	2	1.04	0.1	1	0	0.96	0.04	0.04
Bg5SR42523	192	2	1.09	0.17	0.92	0.08	0.92	0.08	0.08
Bg5SR42999	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR43415	184	2	1.81	0.64	0.98	0.02	0.55	0.45	0.45
Bg5SR43763	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR44141	192	2	1.02	0.06	0.98	0.02	0.98	0.02	0.02
Bg5SR44241	192	2	1.17	0.27	0.99	0.01	0.86	0.14	0.14
Bg5SR44262	192	2	1.01	0.03	0.99	0.01	0.99	0.01	0.01
Bg5SR44446	192	2	2	0.69	0	1	0.5	0.5	0.5
Bg5SR13037	192	2	2	0.69	0	1	0.5	0.5	0.5
Mean	191	1.99	1.46	0.38	0.63	0.37	0.74	0.26	0.26
St. Dev		0.09	0.43	0.28	0.43	0.43	0.21	0.21	0.21

Figures

Figure 1

Analysis of proportion of 45,044 SSRs detected in 334 Mbp bottle gourd available non-redundant RAD genomic DNA sequence data.

Figure 2

Relative frequency (%) of SSR types with different number of repeats, present in the bottle gourd genome.

Figure 3

Genomic microsatellite markers are anchored onto the 11 linkage groups of chromosomes. The lengths of the bars are proportional to their genetic or physical lengths.

Figure 4

DNA profile of 96 genotypes of bottle gourd showing polymorphism with SSR primer SSR21730 (M is 50 bp ladder).

Figure 6

Results from the program STRUCTURE 2.3.4 analysis of bottle gourd samples from different locations in India. Plot displays mean delta k values for 10 independent runs for each value of K from K = 2–9. The highest value was at K = 4, indicating that the 4 most close related germplasm.

Figure 7

Population structure inferred presenting distribution of germplasm into three population groups with some germplasms showing gene flow.

BG Vs Melon (Cm)

BG Vs Cucurbit (Cs)

BG Vs Watermelon (Cl)

BG Vs Cl, Cm and Cs

Figure 8

Genome syntenic relationships of bottle gourd with other cucurbits like cucumber, melon, watermelon. Syntenic relationship of mapped bottle gourd microsatellite markers with cucumber, watermelon, melon showing synteny with 181, 188 and 193 microsatellite markers respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.xls](#)