Mining of Simple Sequence Repeats Loci, Genetic Relationship And Population Structure of Bottle Gourd (*Lagenaria Siceraria* (Molina) Standl.) Accessions With Different Geographical Origin Using Single Nucleotide Polymorphism (SNPs) Markers

Rodrigo Iván Contreras-Soto (contrerasudec@gmail.com)
Universidad de O'Higgins https://orcid.org/0000-0001-6468-9394
Ariel Salvatierra
Centro de Estudios Avanzados en Fruticultura
Carlos Maldonado
Universidad de O'Higgins
Jacob Mashilo
LDA: Limpopo Provincial Government Department of Agriculture and Rural Development

Research Article

Keywords: AMOVA, Lagenaria, Genotyping-by-sequencing, molecular markers

Posted Date: June 18th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-597880/v1

Abstract

Lagenaria siceraria (Molina) Standl. (2n = 2x = 22) is an important horticultural and medicinal crop grown worldwide serving for food and pharmaceutical industries. The crop exhibit extensive phenotypic and genetic variation useful for cultivar obtention targeting economic traits, however limited genomic resources are available for effective germplasm characterization into breeding and conservation strategies. This study determined the genetic relationships and population structure in a collection of different accessions of bottle gourd prevenient from Chile, Asia, and South Africa by using single nucleotide polymorphism (SNPs) markers and mining of simple sequence repeats (SSR) loci derived from genotyping-by-sequencing (GBS) data. The GBS resulted in 12,766 SNPs molecular markers classified as moderate to highly informative with mean polymorphic information content of 0.29. The mean gene diversity of 0.16, indicated low genetic differentiation of the accessions. Analysis of molecular variance revealed lower differentiation between (36%) than within (48%) bottle gourd accessions suggesting that random mating system dominates inbreeding. Population structure revealed two genetically differentiated groups comprising of South Africa accessions and an admixed group with genotypes of Asian and Chilean origin. The results of SSR loci mining from GBS data should be developed and validated before being used in diverse bottle gourd accessions. The SNPs markers developed in the present study are useful genomic resources in bottle gourd breeding programs for assessing the extent of genetic diversity for effective parental selection and breeding.

Introduction

Bottle gourd [*Lagenaria siceraria* (Mol.) Standl., 2n = 2x = 22] or calabash) is a diploid, monoecious, and self-pollinating vegetable crop belonging to the genus *Lagenaria* of the *Cucurbitaceae* family (Achigan-Dako et al. 2008). The crop is used for diverse and beneficial uses including food, feed and medicinal purposes. The fresh and tender fruits are cooked as food and the dry fruits for making containers for food and grain storage, decoration and musical instruments (Jeffrey et al. 1976; Kalpana et al. 2020). The cultivated bottle gourd is also used as rootstock for production of sweet watermelon (*Citrullus lanatus* var. *lanatus*) to control soil-borne diseases, leaf diseases, low soil temperature and improve nitrogen-use efficiency (Yetisir and Sari, 2003; King et al. 2008; Ulas et al. 2019; Aslam et al. 2020) and improve fruit quality (Guler et al. 2013, 2014).

Bottle gourd is thought to be one of the first plant species to be domesticated for human use approximately 10,000 years ago (Decker-Walters and Wilkins-Ellert, 2004; Erickson et al. 2005). Archaeological evidence suggested bottle gourd originated in Africa (Decker-Walters and Wilkins-Ellert, 2004) and comprised of two subspecies namely: the African *L. siceraria* ssp. *siceraria* and the Asian *L. siceraria* ssp. *asiatica* (Kobiakova, 1930; Schlumbaum and Vandorpe, 2012). Although bottle gourd is native of Africa, the species has been widely grown worldwide attributed to its abundant genetic and morphological variation allowing adaptation to diverse growing environments (Erickson et al. 2005; Schlumbaum and Vandorpe, 2012; Mashilo et al. 2017b). The cross-pollinating nature of the crop resulted in phenotypic variation for fruit traits including fruit shape and size (Sivaraj and Pandravada, 2005; Yetişir et al. 2008; Mashilo et al. 2016b), and

seeds morphology (Buthelezi et al. 2019). Fruit and seed characteristics are economic traits for cultivar obtention in this crop targeting various domestic and industrial applications.

The extent of genetic diversity in bottle gourd have been previously assessed employing various molecular markers. In India, Sarao et al. (2014) fingerprinted 20 accessions of bottle gourd using 20 simple sequence repeat (SSR) markers and reported high genetic diversity among accessions. Mashilo et al. (2016a) using 11 SSR markers selected distantly related bottle gourd landraces of South Africa origin. Xu et al. (2014) using 3226 SNPs markers identified two distinct groups among Chinese bottle gourd accessions based on fruit shape rather than collection site. Until now, molecular markers have been used to study population structure and genetic relationships of *L. siceraria*, such as inter-sequence simple repeats (Bhawna et al. 2014), SSRs (Sarao et al. 2014; Mashilo et al. 2016b) and SNPs (Konan et al. 2020; Xu et al. 2014). Next-generation sequencing (NGS)-based SNPs are the most widely used molecular markers to study genome-wide association, population structure, genomic selection, and genetic diversity due to their genome-wide abundance, particularly when a large number of markers are required (Bhattacharjee et al. 2020; Yang et al. 2020). Genotyping-by-sequencing (GBS) has emerged as one NGS-based genotyping platform for marker design and development (Bhattacharjee et al. 2020; Yang et al. 2020), in fact, the NGS technology provide large amounts of sequence data to develop numerous SNP and microsatellite markers at whole genome scale (Zhu et al. 2016). Also, this approach provides accurate results independently of the population or target species. Moreover, GBS can obtain a high marker density without previously available genomic information, with which it can reveal the extent of genetic relatedness and genetic variation within and between cultivated and wild species (Pereira-Dias et al. 2019; Bhattacharjee et al. 2020).

To date there are limited genomic resources developed for bottle gourd germplasm characterization. This has to some extent limited breeding efforts to determine heterotic groups for hybrid development, release, and commercialization of bottle gourd cultivars with desired attributes for farmers, consumers and for food and pharmaceutical industries. Also, quantitative trait loci controlling the expression of key qualitative and quantitative traits remains largely unexplored in bottle gourd partly owing to limited development of genomic resources. In the present study, we developed GBS that resulted in development of 12,766 SNPs molecular markers distributed across 11 chromosomes of bottle gourd. Therefore, the purpose of this study was to determine the genetic relationships and population structure in a collection of different accessions of bottle gourd from Chile, Asia, and South Africa using the new-developed SNPs markers and mining of SSR loci derived from GBS data.

Material And Methods

Plant material

A germplasm collection consisting of 25 bottle gourd accessions originating from different geographic areas of Asia (4), South Africa (15) and South America (6) was used for the current study. Fifteen bottle gourd genotypes of South Africa were local varieties grown by farmers in the Limpopo Province of South Africa and sourced from the Limpopo Department of Agriculture and Rural Development (Towoomba Research Station), South Africa. The four accessions of Asia were sourced from the Genetic Resource

Center of Japan, specifically from the National Agriculture and Food Research Organization (NARO), whereas the accessions of South America were collected and comprised of local populations of Chile and Brazil. Details of the accessions are presented as supplementary material (Table S1).

GBS sequencing, reads clustering and SNP calling

Genomic DNA of the 25 accessions was extracted from young leaves collected from three-weeks year-old seedlings by using the QIAGEN DNeasy Plant Mini Kit for DNA extraction (QIAGEN; https://www.qiagen.com) following the manufacturer's instructions. We evaluated the quality of DNA via agarose gel electrophoresis and measured the fluorometric quantification by Qubit 2.0 and Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific; https://www.thermofisher.com/). The genotyping-by-sequencing data was generated following the Elshire et al. (2011) method and included the following changes: 100 ng of genomic DNA and 3.6 ng of total adapters were used, the genomic DNAs were restricted with ApeKI enzyme and the library was amplified with 18 PCR cycles. After PCR, the pooled products were purified and quantified for sequencing on the Illumina HiSeq 2000 flow cell for sequencing.

Reads and tags (fastq) found in each sequencing lane from 96 barcodes produced a total read pairs of 485 million of reads and an average of 18.5 million of high-quality read pair count. The reads for both ends of the pair-end data were combined into individual per-sample files, and aligned to the bottle gourd inbred line USVL1VR-Ls reference genome using bowtie2 (Wu et al. 2017). The preset –sensitive, end-to-end mapping parameters were used, and the sorted alignments were subsequently used for SNP calling using the Stacks 2.5 pipeline (http://catchenlab.life.illinois.edu/stacks/). Alignment and merging resulted in a total of 71,212 called SNPs.

After removing lines with failed data, the GBS data from the 25 accessions were stored in Variant Call Format version 4.1 (Danecek et al. 2011). Genotyping-by-sequencing datasets typically have high rates of missing data (Poland and Rife, 2012). The linkage disequilibrium k nearest neighbor imputation (Money et al. 2015) method was used to impute missing values in this dataset. Only SNPs with a minor allele frequency > 0.05 and < 25% missing data were filtered, resulting in 12,766 high-quality polymorphic SNPs. The SNP calling was performed using TASSEL version 5.2 in the GBS pipeline (Glaubitz et al. 2014).

Analysis of genetic diversity parameters and molecular variance

Genetic diversity of 25 bottle gourd accessions was analyzed with 12,766 SNPs markers by using the poppr package of the R-software (Kamvar et al. 2014). The filtered SNPs were used to calculate the genetic diversity parameters such as minor allele frequency (MAF), polymorphic information content (PIC), expected heterozygosity (He), and observed heterozygosity (Ho). These analyses were carried out in R-package. The PIC value of an l-allele locus can be calculated as:

$$\mathrm{PIC} = 1 - \sum_{i=1}^{l} P_i^2 - \sum_{i=1}^{l-1} \sum_{j=i+1}^{l} 2P_i^2 P_j^2$$

where Pi and Pj are the population frequency of the ith and jth allele.

Analysis of molecular variance (AMOVA) was carried out by using the poppr package in R to detect population differentiation (Excoffier et al. 1992). Transitions/transversions and percentage of heterozygous positions were determined using SNiPlay3 (Dereeper et al. 2011).

Population structure and genetic relationship

The genetic relationship among the landraces of bottle gourd was calculated based on identity-by-state (IBS) distance that represent a kinship matrix, using the software TASSEL 5.2 (Bradbury et al. 2007). The population structure was inferred with the Markov Chain Monte Carlo (MCMC) algorithm for the generalized Bayesian clustering method implemented in the Structure software (Pritchard et al. 2000). Consequently, 10 independent runs of MCMC sampling were implemented for numbers of groups (K parameter), varying from 2 to 5. For each run, the initial burn-in period was set to 10,000 with 110,000 MCMC iterations, under the non-admixture model, and with prior information on the individual's origin. The optimal value of K was estimated from the second-order change rate of the probability function with respect to K (ΔK), as proposed by Evanno et al. (2005).

Mining of simple sequence repeats markers

The Illumina raw reads data were preprocessed to generate clean reads and then analyzed using the core Stack pipeline of Stacks v.2.5 software with default parameters. Each consensus sequence resulting from the Stack pipeline was then screened for simple sequence repeats (SSRs) using GMATA package with default parameters (Wang and Wang, 2016). The acquired SSRs were considered to only represent those containing perfect repeats of SSRs whose basic motifs ranged from 2 to 6 bp with defined minimum repeat units of five iterations for di-, tri-, tetra-, penta-, hexa- and heptanucleotide repeats.

Results

GBS Analysis

Genome sequencing of the 25 bottle gourd accessions using GBS generated a total of 485 million reads pairs, with an average read pair count of 18.5 million. Each of the 25 sample reads was mapped to '*Lagenaria siceraria* var. USVL1VR-Ls'. In the GBS analysis a total of 71,212 called and unfiltered SNPs were detected as raw SNP markers. Of these, 12,766 filtered SNPs were obtained and distributed across the eleven chromosomes *of L. siceraria*. The numbers of homozygote and heterozygote SNP loci ranged from 9,865 (CLS-013) to 10,594 (CLS-024) with an average of 10,194 and from 2334 (CLS-024) to 3063 (CLS-013) with an average of 2734, respectively (Table 1). The average homozygote rate was approximately 78.9%, and the average heterozygote rate was 21.1% (Fig. 1). Transversion SNPs (62%, 37790 SNPs) were more frequent than transition (38%, 23141 SNPs). Of these, the C/G transversion (38.6%) accounted for the highest frequency, whereas C/T transitions (19.2%) occurred at the lowest frequency among all the 60,931 SNPs (Fig. 2).

The average PIC value across all the markers and chromosomes was 0.26, whereas the observed heterozygosity ranged from 0.15 to 0.22 with an average of 0.18. The expected heterozygosity ranged

between 0.15 and 0.16, with a mean of 0.16. Minor allele frequency (MAF) ranged between 0.21 and 0.242, with an average of 0.23. The highest PIC and MAF were on chromosome ten, whereas the lowest were on chromosome eight (Table 1).

Table 1
Summary statistics of genetic diversity parameters generated by single nucleotide polymorphic markers across eleven chromosomes of *L. siceraria*.

Chromosome number	Number of SNPs markers	PIC	MAF	Ho	He
1	1285	0.298	0.236	0.184	0.159
2	1580	0.297	0.238	0.208	0.161
3	1305	0.305	0.239	0.179	0.162
4	1448	0.287	0.224	0.186	0.155
5	1272	0.292	0.228	0.177	0,157
6	1044	0.291	0.225	0.166	0.156
7	855	0.286	0.222	0.176	0.154
8	1103	0.275	0.211	0.150	0.147
9	1059	0.297	0.234	0.190	0.159
10	824	0.302	0.242	0.211	0.163
11	991	0.295	0.231	0.172	0.158
Total/Average	12766	0.293	0.230	0.183	0.157

PIC: Polymorphic Information Content; MAF: Minor allele frequency; H_O: observed heterozygosity; H_E: expected heterozygosity

AMOVA

According to AMOVA, the hypothesis of random mating between the three bottle gourd populations that represent the geographical origin (Asia, South Africa, and South America) was rejected, with strong evidence that these populations are significantly differentiated at all stratifications (Table 2).

According to the phi-statistics, there was relatively high differentiation between the different levels of comparison. The lowest differentiation was reported among samples within the same population or geographical origin (25%). Substantial differentiation between populations was reported (36%). However, 52% of the differentiation occurred within samples (Table 2).

Table 2
Analysis of molecular variance among bottle gourd accessions showing percentage of molecular variance explained by each source of variation.

Component of differentiation	DF	Mean Square	PVE (%)	Phi-statistics
Between populations	2	28618	35.9	PT = 0.36
Between samples within populations	22	3958	16.3	SP = 0.25
Within samples	25	2357	47.8	ST = 0.52
Total	-	-	100	

DF: degree of freedom; PVE: percentage of variance explained, PT = population-total differentiation; SP = sample-population differentiation, ST = sample-total differentiation

Population structure

Population structure and genetic relationship analysis revealed two genetically differentiated groups (Figs. 3 and 4). Table 3 shows the results of the statistical parameters that define the number of groups or populations that represented the population structure of the 25 accessions of bottle gourd. Specifically, in Fig. 3, the blue color represents the percentage of membership of the sixteen accessions of South African origin and the orange color represents the percentage of membership of the nine genotypes of Asian and South American origin. The Fig. 4 showed the heatmap of the kinship among the 25 accessions of *L. siceraria*. The red and orange color represented the higher relationship within bottle gourd accessions, whereas the yellow color corresponds to lower relationship.

Table 3
Number of individuals by population, and statistics parameters and credible intervals for each cluster of 25 accessions of *L. siceraria.*

Cluster	Individuals	Mean	Median	Mode *	SD	95% Credible Interval	
						Lower	Upper
I	9	0.40	0.40	0.40	0.0062	0.39	0.41
II	16	0.62	0.61	0.61	0.0052	0.60	0.62

* Kernel density estimates of the mode from marginal posterior distributions.

Bottle gourd SSR locus identification and the frequency of SSRs

We used high-quality read pair count sequences derived from GBS data to identify SSR loci in a collection of 25 bottle gourd accessions. The search for SSR-containing regions was restricted to motif of di-, tri-, tetra-, penta-, hexa-, and heptanucleotides. A total of 95,635 SSRs loci with di-, tri-, tetra-, penta-, hexa- and heptanucleotide repeats of five or more repeats were identified from the GBS data. These SSR loci consisted of 69,682 dinucleotide repeats (72.86%), 21,641 trinucleotide repeats (22.63%), 3,203 tetranucleotides repeats (3.35%), 599 pentanucleotide repeats (0.63%), 356 hexanucleotide repeats (0.37%) and 154 heptanucleotides repeats (0.16%) (Table 4). Dinucleotides and trinucleotides were identified as the

most abundant SSR class, representing the 95,49% of the SSR motif classes. The repeat motif AT/AT (26,274) was the most frequent into the dinucleotide SSR, representing 37.71% of the total dinucleotides, and the repeat motif AAT/ATT (7,592) was the most frequent into the trinucleotide SSR, representing 31,08% of the trinucleotides (Fig. 5).

Table 4
Summary of bottle gourd SSRs identified based on GBS sequences.

SSR motifs	Number of repeat units of each SSR motif								Frequency (%)
	5	6	7	8	9	10	>10	Total	
Dinucleotide	28400	10526	6872	4695	3539	2679	12971	69,682	72.86
Trinucleotide	9670	4589	2399	1386	855	603	2139	21,641	22.63
Tetranucleotide	2372	537	160	68	28	12	26	3203	3.35
Pentanucleotide	424	125	33	10	1	3	3	599	0.63
Hexanucleotide	250	82	16	4	2	1	1	356	0.37
Heptanucleotide	105	20	15	2	1	2	9	154	0.16
Total	41221	15879	9495	6165	4426	3300	15149	95635	100.00

Discussion

Effective use of bottle gourd genetic resources for cultivar obtention and for conservation requires development of genomic tools for marker-assisted breeding. During the last decade, significant progress has been made in the development of genomic resources in bottle gourd. Most of these genomic resources provide valuable information about genetic relationships among genotypes for effective selection and use in breeding programs (Xu et al. 2014; Wu et al. 2017; Wang et al. 2018). Despite significant progress, there are generally very limited genomic resources developed for bottle gourd limiting breeding efforts to develop competitive genotypes for agricultural production and in the nutraceutical and pharmaceutical industries. The present study identified SNPs molecular markers distributed across 11 chromosomes of bottle gourd employing genotyping-by-sequencing platform which were then used to determine genetic relationships and population structure in a collection of bottle gourd accessions of African, Asian, and South American origins, and subsequently identified SSR loci from the GBS sequences. Among the high throughput sequencing technologies, GBS is considered the most cost-effective tool to identify and genotype a large number of polymorphisms at genome-scale (Wu et al. 2017). Here, we used Elshire-GBS method and *Lagenaria siceraria* var. USVL1VR-Ls as reference genome which resulted in a set of 12,766 filtered SNPs markers. A recent GBS study used to confirm the varietal status of bottle gourd accessions produced 22,575 SNPs (Konan et al. 2020), which was higher than the present study. Others high throughput studies conducted in bottle gourd used the Restriction site-associated DNA sequencing (RAD-Seq), a form of GBS that generate low coverage genome sequencing in which reference genomes are not available (Xu et al. 2014; Wu et al. 2017). In addition, Wu et al. (2017) using RAD-Seq and aligning to the Hangzhou gourd

reference genome detected 19,226 SNPs, similar with the present findings. On the contrary, Xu et al. (2014) using RAD-Seq genotyping identified 3,226 SNPs and Xu et al. (2011) using partial sequencing only discovered 3,913 putative SNPs. These differences between the current study and previous results may be due to high read depth variation of RAD-Seq or the high levels of missing data of Elshire-GBS (Scheben et al. 2017) and the average coverage which typically varies between these reduced-representation sequencing methods. For instance, while RAD-seq involves sequencing fragments to moderate coverage between 5x and 15x (Fountain et al. 2016), Elshire-GBS studies tend to reach low coverage of ~ 1x (Swarts et al. 2014). Despite these differences, the generated SNPs markers and SSR loci are a useful genomic resource for genetic analysis and breeding in bottle gourd for diverse applications, however, in subsequent studies, a final set of SSR loci should be developed and validated before being used in diverse bottle gourd accessions collected from different regions of the world.

For instance, in this study, the most abundant class of SSRs identified from GBS sequences was comprised by dinucleotide and trinucleotide repeats. Similar results have been reported previously for bottle gourd. Xu et al. (2011), for example, identified that dinucleotide and trinucleotide repeats were the most abundant, while mononucleotide and pentanucleotide repeats were relatively rare. Moreover, the high frequency of dinucleotide and trinucleotide repeats is consistent with other cucurbit species, including cucumber and watermelon (Ren et al. 2009; Zhu et al. 2016b). Furthermore, similar to our results, the AT-rich motifs have been the predominant motif in all nucleotide repeats in melon, watermelon, cucumber, and bottle gourd genomes (Ren et al. 2009; Zhu et al. 2016a; Zhu et al. 2016b).

In a breeding program, the extent of genetic diversity and population relationships among the germplasm is useful to identify distantly related parents for hybridization to develop genetically improved genotypes of bottle gourd for rootstocks, food, feed and medicinal purposes. For this reason, in different regions, several studies have been conducted to determine the genetic diversity of bottle gourd accessions (Gürcan et al. 2015; Mashilo et al. 2016b; Ibrahim, 2021). In this study, the accessions of bottle gourd were collected from Chile, Japan (Philippines, South Korea), and South Africa. Most of the Asian accessions share similar genetic background to South African accessions which been previously assayed using SSR markers (Mashilo et al. 2017a). In the current study, various genetic parameters were estimated using SNPs markers including Ho, He and PIC values with mean values of 0.18, 0.16 and 0.29, respectively. Gürcan et al. (2015), genotyped thirty-one bottle gourd accessions from USA, India, Nigeria and Russia using SSR markers and reported mean values of 0.50, 0.13, and 0.50 for He, Ho and PIC, in that order. Also, Mashilo et al. (2016b) using SSR markers reported high average values for He = 0.657 and PIC = 0.57 among bottle gourd accessions, higher than values reported in the present study. Botstein et al. (1980) classified the PIC values in to three categories (1) if the PIC value of the marker is more than 0.5, the marker is considered a highly informative, (2) if the PIC value ranged from 0.25 to 0.5, the marker is a moderately informative, and (3) if the PIC value less than 0.25, then the marker is slightly informative. Based on Botstein classification, SNPs markers generated in the present study are moderately informative. A recent study indicated that PIC values calculated with SNPs markers showed lowest values compared to SSR markers (Singh et al. 2013; Liu et al. 2017). This can be attributed to the bi-allelic nature of the SNPs which is restricted to PIC values ranging

from 0.0 to 0.5 (i.e., when the two alleles have identical frequencies), whereas for SSR markers which are multi-allelic PIC value can vary between 0.5 and 1.0 (Singh et al. 2013; Eltaher et al. 2018).

Expected heterozygosity is usually preferred to assess genetic diversity, because it is less sensitive to the sample size than the observed heterozygosity (Chesnokov and Artemyeva, 2015). According to Chesnokov and Artemyeva (2015), when H_o and H_e are similar (i.e., not significantly different), the crossing in the population is almost accidental. When $H_o < H_e$, it is an inbred population, and when $H_o > H_e$, the random mating system dominates inbreeding in the population. Our results showed that H_o was slightly higher than H_e, suggesting that random mating system dominates inbreeding in the assessed bottle gourd germplasm. Moreover, population differentiation indicated a higher variation within sample, a common characteristic of cross-pollinated plants which can reduce the loss of genetic diversity through large gene flow. As proposed by Mashilo et al. (2016b), this could be attributed to the high out-crossing nature of bottle gourd or long-term selection of the crop by farmers for diverse uses.

Population structure and genetic relatedness are useful to understand genetic diversity, differentiate the population according to their geographical origin and conduct association mapping studies. Based on population structure analysis, two genetically differentiated groups were identified; the first including all the accessions originated from South Africa and the second group comprising of Asian and Chilean accessions. These results agree with previous studies conducted in bottle gourd, which reported that clustering of different landraces was independent of geographical location (Yetişir et al. 2008; Sarao et al. 2014; Gürcan et al. 2015; Mashilo et al. 2016b). Another explanation is that founder effect followed by artificial selection based on fruit shape which tend to generate high genetic similarity (Xu et al. 2011; Yildiz et al. 2015). In crop improvement programs, germplasm collection missions should be based on morphological variation rather than geographical origin (Mashilo et al. 2016b). Heiser (1973) classified bottle gourd into two subspecies: Asian and American-African subspecies. These authors postulated that African wild bottle gourd floated to the shores of America and were independently domesticated there. Using various molecular markers, different results on the phenomenon have been reported. For example, Erickson et al. (2005) using SNP markers within chloroplast DNA concluded that American bottle gourds were more closely related to Asian than to African gourds, whereas Decker-Walters and Wilkins-Ellert, (2004) by using RAPD molecular markers revealed that American germplasm is distinct and primarily originated from Africa but possesses Asian genetic profiles. Similar with Erickson et al. (2005), our results supported the idea that one group is only composed by genotypes of South Africa, and the other correspond to an admixture group with genotypes from Asia and South America.

Conclusions

The present study genotyped bottle gourd accessions of diverse origins using new-developed single nucleotide polymorphism markers. A total of 12,766 SNPs molecular markers were generated using genotyping-by-sequencing which were classified as moderate to highly informative. Low genetic differentiation was observed among the assessed bottle gourd accessions using SNPs markers. Random mating system was found to dominate inbreeding in the assayed bottle gourd population. Accordingly, two

genetically differentiated groups comprising of South African accessions and an admixed group with genotypes of Asian and Chilean origin were identified. The results of SSR loci mining from GBS data should be developed and validated before being used in diverse bottle gourd accessions. The SNPs developed in the present study are a useful genomic resource for bottle gourd breeding targeting development of genetically improved genotypes for diverse uses including rootstocks, food, feed and medicine.

Declarations

Acknowledgments: RCS thanks to CEAF (Centro de Estudios Avanzados en Fruticultura) for the support of the project, National Agricultural and food Research Organization (NARO) and Limpopo Department of Agriculture and Rural Development of South Africa for the support of plant material.

Author's contributions: Conceptualization, Methodology, Resources: RCS; Writing draft & editing: RCS, AS, CM and JM; Data analysis: RCS and CM; Experimental work: RCS and AS. All authors reviewed and approved the manuscript.

Authors' information (optional): Not Applicable

Availability of data and materials: Datasets supporting the conclusions of this article are included within the article.

Code availability: Not applicable.

Funding: This work was supported by the National Commission for Scientific and Technological Research (CONICYT, Chile) FONDECYT (grant number 11180278).

Ethics approval: Not applicable.

Consent to participate: Not applicable.

Consent for publication: Not applicable.

Competing interests/Competing interests: The authors declared no conflict of interest.

References

1. Achigan-Dako EG, Fuchs J, Ahanchede A, Blattner FR (2008) Flow cytometric analysis in Lagenaria siceraria (Cucurbitaceae) indicates correlation of genome size with usage types and growing elevation. Plant Syst Evol 276(1–2):9–19. doi:10.1007/s00606-008-0075-2
2. Aslam W, Noor RS, Hussain F, Ameen M, Ullah S et al (2020) Evaluating morphological growth, yield, and postharvest fruit quality of cucumber (Cucumis sativus L.) grafted on cucurbitaceous rootstocks. Agric 10(4). doi:10.3390/agriculture10040101
3. Bhattacharjee R, Agre P, Bauchet G, de Koeyer D, Lopez-Montes A et al (2020) Genotyping-by-sequencing to unlock genetic diversity and population structure in white yam (dioscorea rotundata

poir.). Agronomy 10(9). doi:10.3390/AGRONOMY10091437

4. Bhawna M, Abdin Z, Arya L, Saha D, Sureja AK et al (2014) Population structure and genetic diversity in bottle gourd [Lagenaria siceraria (Mol.) Standl.] germplasm from India assessed by ISSR markers. Plant Syst Evol 300(4):767–773. doi:10.1007/s00606-014-1000-5
5. Botstein D, White RL, Skolnick M, Davis RW (1980) Botstein. Am J Hum Gen 32: 314–331. papers2://publication/uuid/0B80518E-A22B-41F3-BE43-171F51007E42
6. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y et al (2007) TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23(19):2633–2635. doi:10.1093/bioinformatics/btm308
7. Buthelezi LG, Mavengahama S, Ntuli NR (2019) Morphological variation and heritability studies of lagenaria siceraria landraces from northern Kwazulu-Natal, South Africa. Biodiversitas 20(3):922–930. doi:10.13057/biodiv/d200342
8. Danecek P, Auton A, Abecasis G, Albers CA, Banks E et al (2011) The variant call format and VCFtools. Bioinformatics 27(15):2156–2158. doi:10.1093/bioinformatics/btr330
9. Decker-Walters D, Wilkins-Ellert M (2004) Discovery and genetic assessment of wild bottle gourd from Zimbabwe. Econ. Bot. (1995). http://link.springer.com/article/10.1663/0013-0001(2004)058%5B0501:DAGAOW%5D2.0.CO%3B2
10. Dereeper A, Nicolas S, Le Cunff L, Bacilieri R, Doligez A et al (2011) SNiPlay: A web-based tool for detection, management and analysis of SNPs. Application to grapevine diversity projects. BMC Bioinformatics 12:1–14. doi:10.1186/1471-2105-12-134
11. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6(5). doi:10.1371/journal.pone.0019379
12. Eltaher S, Sallam A, Belamkar V, Emara HA, Nower AA et al (2018) Genetic diversity and population structure of F3:6 Nebraska Winter wheat genotypes using genotyping-by-sequencing. Front. Genet. 9(MAR). doi: 10.3389/fgene.2018.00076
13. Erickson DL, Smith BD, Clarke AC, Sandweiss DH, Tuross N (2005) An Asian origin for a 10,000-year-old domesticated plant in the Americas. Proc. Natl. Acad. Sci. U. S. A. 102(51): 18315–18320. doi: 10.1073/pnas.0509279102
14. Evanno RS, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol: 2611–2620. doi:10.1111/j.1365-294X.2005.02553.x
15. Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics. doi:10.5962/bhl.title.86657
16. Fountain ED, Pauli JN, Reid BN, Palsbøll PJ, Peery MZ (2016) Finding the right coverage: the impact of coverage and sequence quality on single nucleotide polymorphism genotyping error rates. Mol Ecol Resour 16(4):966–978. doi:10.1111/1755-0998.12519

17. Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ et al (2014) TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS One 9(2). doi:10.1371/journal.pone.0090346
18. Guler Z, Candir E, Yetisir H, Karaca F, Solmaz I (2014) Volatile organic compounds in watermelon (Citrullus lanatus) grafted onto 21 local and two commercial bottle gourd (Lagenaria siceraria) rootstocks. J Hortic Sci Biotechnol 89(4):448–452. doi:10.1080/14620316.2014.11513105
19. Guler Z, Karaca F, Yetisir H (2013) Volatile compounds in the peel and flesh of cucumber (cucumis sativus L.) grafted onto bottle gourd (Lagenaria siceraria) rootstocks. J Hortic Sci Biotechnol 88(2):123–128. doi:10.1080/14620316.2013.11512945
20. Gürcan K, Say A, Yetişir H, Denli H (2015) A study of genetic diversity in bottle gourd [Lagenaria siceraria (Molina) Standl.] population, and implication for the historical origins on bottle gourds in Turkey. Genet Resour Crop Evol 62(3):321–333. doi:10.1007/s10722-015-0224-8
21. Heiser CB Jr (1973) in Tropical Forest Ecosystems in Africa and South America: A Comparative Review, eds. Meggers, B, Ayensu E, Duckworth W (Smithsonian Institution Press, Washington, DC), pp 121–128
22. Ibrahim EA (2021) Genetic diversity in Egyptian bottle gourd genotypes based on ISSR markers. Ecol Genet Genomics 18(January):6–11. doi:10.1016/j.egg.2021.100079
23. Kalpana VN, Alarjani KM, Rajeswari VD (2020) Enhancing malaria control using Lagenaria siceraria and its mediated zinc oxide nanoparticles against the vector Anopheles stephensi and its parasite Plasmodium falciparum. Sci Rep 10(1):1–12. doi:10.1038/s41598-020-77854-w
24. Kamvar ZN, Tabima JF, Gr̆unwald NJ (2014) Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ. doi:10.7717/peerj.281
25. King SR, Davis AR, Liu W, Levi A (2008) Grafting for disease resistance. HortScience 43(6):1673–1676. doi:10.21273/hortsci.43.6.1673
26. Kobiakova JA (1930) The bottle gourd. Bull Appl Bot Genet Plant Breed 23:475–520
27. Liu Z, Li J, Fan X, Htwe NMPS, Wang S et al (2017) Assessing the numbers of SNPs needed to establish molecular IDs and characterize the genetic diversity of soybean cultivars derived from Tokachi nagaha. Crop J 5(4):326–336. doi:10.1016/j.cj.2016.11.001
28. Mashilo J, Odindo AO, Shimelis HA, Musenge P, Tesfay SZ et al (2017a) Drought tolerance of selected bottle gourd [Lagenaria siceraria (Molina) Standl.] landraces assessed by leaf gas exchange and photosynthetic efficiency. Plant Physiol Biochem 120:75–87. doi:10.1016/j.plaphy.2017.09.022
29. Mashilo J, Shimelis H, Odindo A (2016a) Genetic diversity of bottle gourd (Lagenaria siceraria (Molina) Standl.) landraces of South Africa assessed by morphological traits and simple sequence repeat markers. South African J Plant Soil 33(2):113–124. doi:10.1080/02571862.2015.1090024
30. Mashilo J, Shimelis H, Odindo A (2017b) Phenotypic and genotypic characterization of bottle gourd [Lagenaria siceraria (Molina) Standl.] and implications for breeding: A Review. Sci Hortic (Amsterdam) 222(November):136–144. doi:10.1016/j.scienta.2017.05.020
31. Mashilo J, Shimelis H, Odindo A, Amelework B (2016b) Genetic diversity of South African bottle gourd [Lagenaria siceraria (Molina) standl.] landraces revealed by simple sequence repeat markers. HortScience 51(2):120–126. doi:10.21273/hortsci.51.2.120

32. Money D, Gardner K, Migicovsky Z, Schwaninger H, Zhong GY et al (2015) LinkImpute: Fast and accurate genotype imputation for nonmodel organisms. G3 Genes. Genomes Genet 5(11):2383–2390. doi:10.1534/g3.115.021667
33. Pereira-Dias L, Vilanova S, Fita A, Prohens J, Rodríguez-Burruezo A (2019) Genetic diversity, population structure, and relationships in a collection of pepper (Capsicum spp.) landraces from the Spanish centre of diversity revealed by genotyping-by-sequencing (GBS). Hortic Res 6(1). doi:10.1038/s41438-019-0132-8
34. Poland JA, Rife TW (2012) Genotyping-by-Sequencing for Plant Breeding and Genetics. Plant Genome 5(3). doi:10.3835/plantgenome2012.05.0005
35. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics
36. Ren Y, Zhang Z, Liu L, Staub JE, Han Y et al (2009) An integrated genetic and cytogenetic map of the cucumber genome. PLoS ONE 4:e5795
37. Sarao NK, Pathak M, Kaur N, Kaur K (2014) Microsatellite-based DNA fingerprinting and genetic diversity of bottle gourd genotypes. Plant Genet Resour Characterisation Util 12(1):156–159. doi:10.1017/S1479262113000385
38. Scheben A, Batley J, Edwards D (2017) Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. Plant Biotechnol J 15(2):149–161. doi:10.1111/pbi.12645
39. Schlumbaum A, Vandorpe P (2012) A short history of Lagenaria siceraria (bottle gourd) in the Roman provinces: Morphotypes and archaeogenetics. Veg Hist Archaeobot 21(6):499–509. doi:10.1007/s00334-011-0343-x
40. Singh N, Choudhury DR, Singh AK, Kumar S, Srinivasan K et al (2013) Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. PLoS One 8(12):1–14. doi:10.1371/journal.pone.0084136
41. Sivaraj N, Pandravada SR (2005) Morphological diversity for fruit characters in bottle gourd germplasm from tribal pockets of Telangana region of Andhra Pradesh, India. Asian Agrihist 9(4):305–310
42. Swarts K, Li H, Romero Navarro JA, An D, Romay MC et al (2014) Novel Methods to Optimize Genotypic Imputation for Low-Coverage, Next-Generation Sequence Data in Crop Plants. Plant Genome 7(3):1–12. doi:10.3835/plantgenome2014.05.0023
43. Ulas A, Doganci E, Ulas F, Yetisir H (2019) Root-growth characteristics contributing to genotypic variation in nitrogen efficiency of bottle gourd and rootstock potential for watermelon. Plants 8(3). doi:10.3390/plants8030077
44. Wang Y, Xu P, Wu X, Wu X, Wang B et al (2018) GourdBase: A genome-centered multi-omics database for the bottle gourd (Lagenaria siceraria), an economically important cucurbit crop. Sci Rep 8(1):1–8. doi:10.1038/s41598-018-22007-3
45. Wang X, Wang L (2016) GMATA: An integrated software package for genome-scale SSR mining, marker development and viewing. Front Plant Sci 7. https://doi.org/10.3389/fpls.2016.01350

46. Wu S, Shamimuzzaman M, Sun H, Salse J, Sui X et al (2017) The bottle gourd genome provides insights into Cucurbitaceae evolution and facilitates mapping of a Papaya ring-spot virus resistance locus. Plant J 92(5):963–975. doi:10.1111/tpj.13722
47. Xu P, Wu X, Luo J, Wang B, Liu Y et al (2011) Partial sequencing of the bottle gourd genome reveals markers useful for phylogenetic analysis and breeding. BMC Genom 12. doi:10.1186/1471-2164-12-467
48. Xu P, Xu S, Wu X, Tao Y, Wang B et al (2014) Population genomic analyses from low-coverage RAD-Seq data: A case study on the non-model cucurbit bottle gourd. Plant J 77(3):430–442. doi:10.1111/tpj.12370
49. Yang X, Tan B, Liu H, Zhu W, Xu L et al (2020) Genetic Diversity and Population Structure of Asian and European Common Wheat Accessions Based on Genotyping-By-Sequencing. Front Genet 11(September):1–14. doi:10.3389/fgene.2020.580782
50. Yetişir H, Şakar M, Serçe S (2008) Collection and morphological characterization of Lagenaria siceraria germplasm from the Mediterranean region of Turkey. Genet Resour Crop Evol 55(8):1257–1266. doi:10.1007/s10722-008-9325-y
51. Yetisir H, Sari N (2003) Effect of different rootstock on plant growth, yield and quality of watermelon. Aust J Exp Agric 43(10):1269–1274. doi:10.1071/EA02095
52. Yildiz M, Cuevas HE, Sensoy S, Erdinc C, Baloch FS (2015) Transferability of Cucurbita SSR markers for genetic diversity assessment of Turkish bottle gourd (Lagenaria siceraria) genetic resources. Biochem Syst Ecol 59:45–53. doi:10.1016/j.bse.2015.01.006
53. Zhu H, Song P, Koo DH, Guo L, Li Y, Sun S, Weng Y, Yang L (2016a) Genome wide characterization of simple sequence repeats in watermelon genome and their application in comparative mapping and genetic diversity analysis. BMC Genom 17:557
54. Zhu H, Guo L, Song P, Luan F, Hu J, Sun X, Yang L (2016b) Development of genome-wide SSR markers in melon with their cross-species transferability analysis and utilization in genetic diversity study. Mol Breeding 36(11):153

Figures

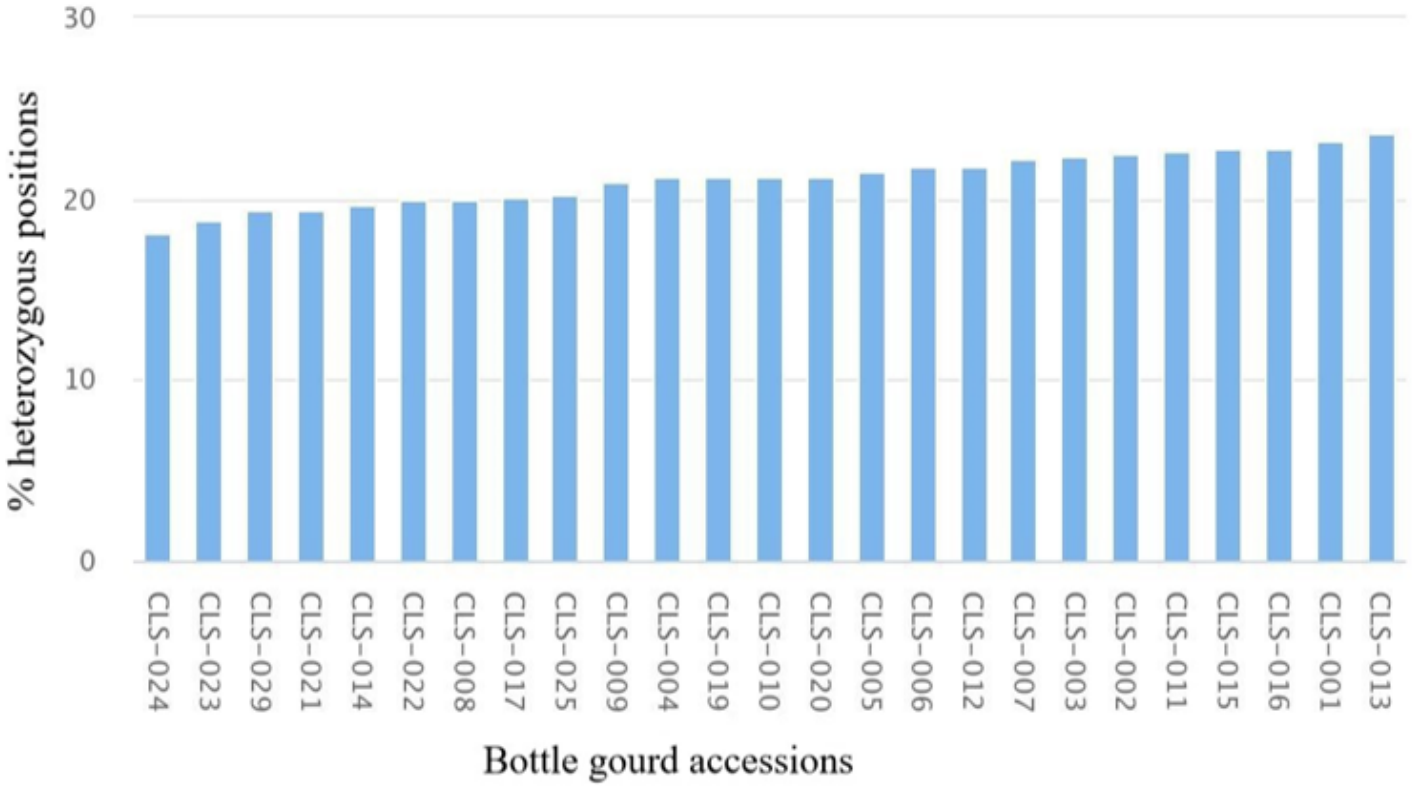

Figure 1

Percentage of heterozygous positions of 25 L. siceraria accessions of diverse geographical origins generated using single nucleotide markers developed using genotyping-by-sequencing

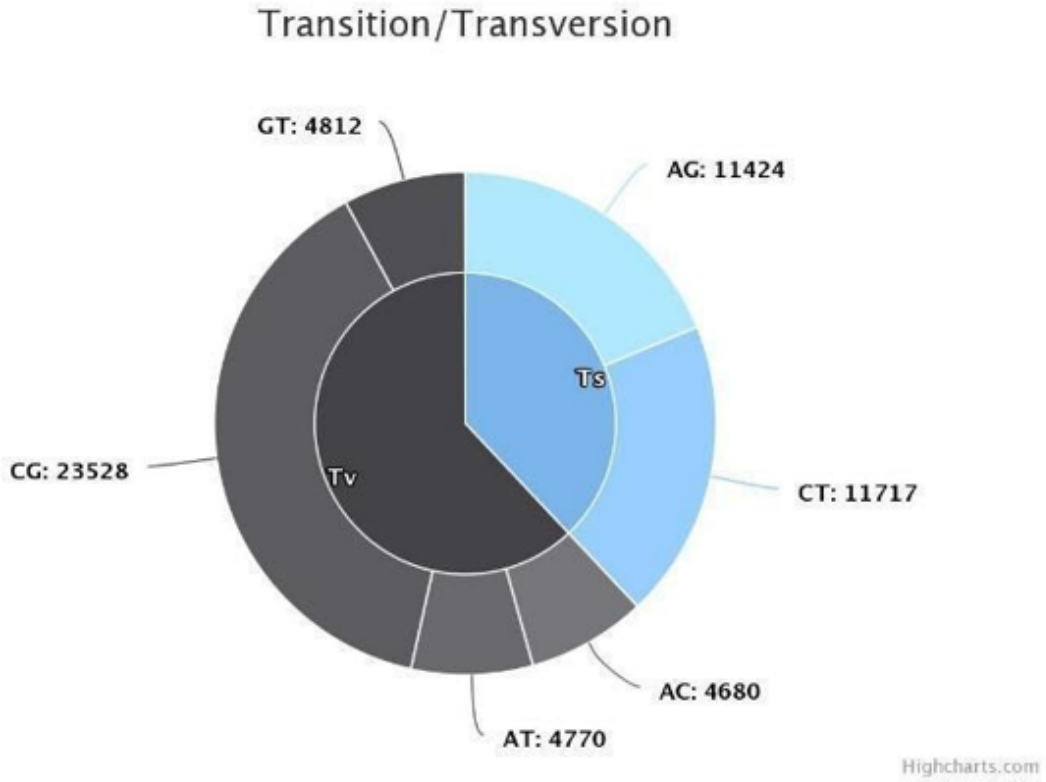

Figure 2

Classification of SNPs markers based on their nucleotide substitutions, either transitions (Ts) or transversions (Tv)

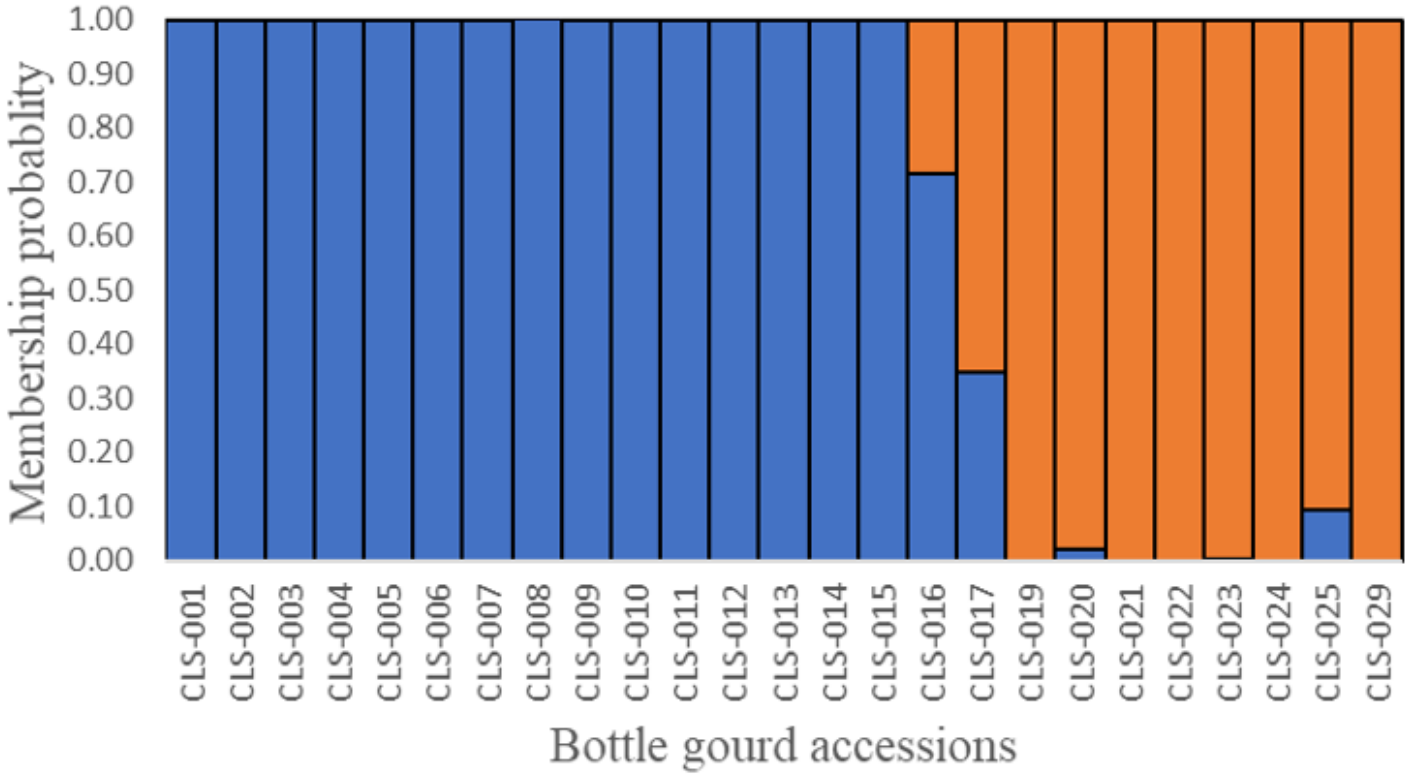

Figure 3

Bar plot of the estimated population structure of 25 L. siceraria accessions (k=2). The vertical bars of the y-axis represented the estimated membership probability of each population to the subgroups. The blue and orange color represent different subgroups, and the x-axis is the bottle gourd accession codes. Populations with colored segments indicates admixed origin

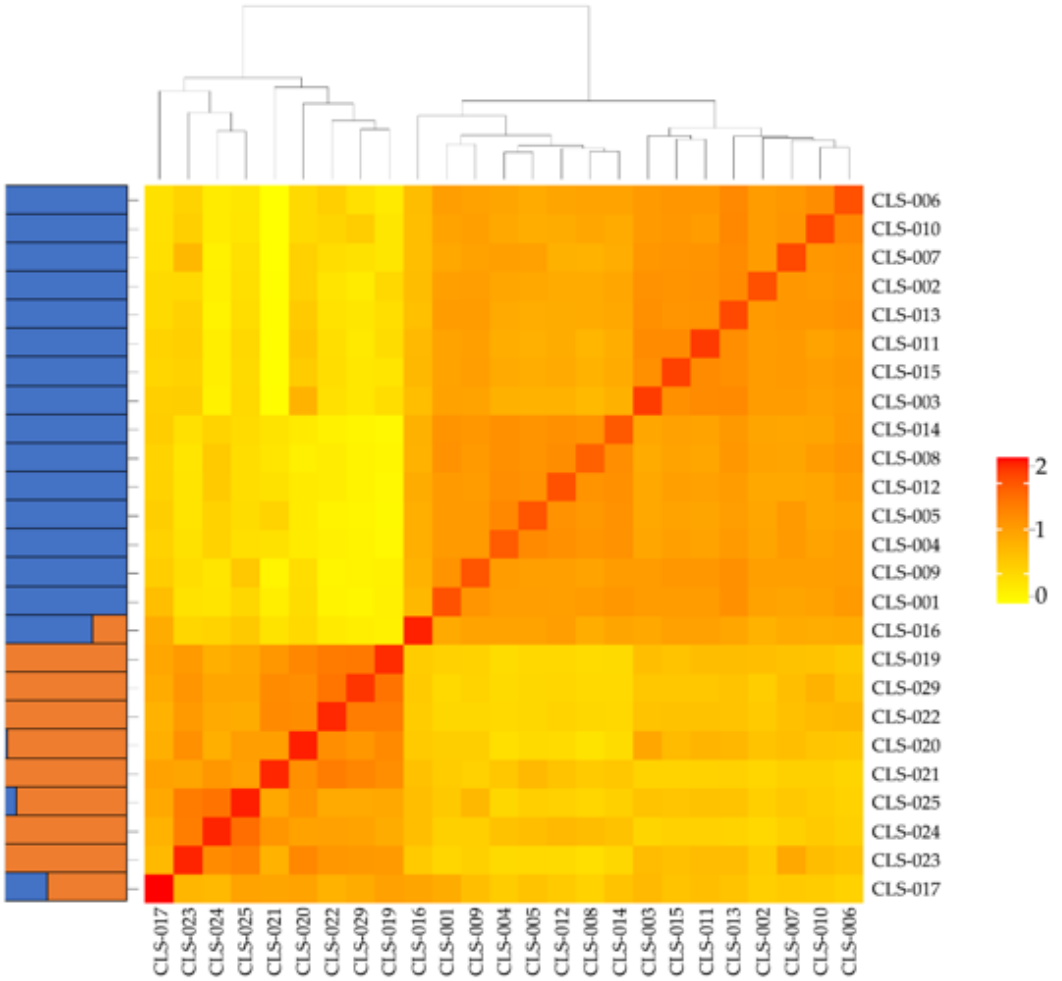

Figure 4

Kinship heatmap showing the genetic relationship among the 25 accessions of L. siceraria based on 71,212 SNPs markers with two main groups. One group is composed by genotypes of South African geographical origin, and the other corresponds to an admixture group with genotypes of different geographical origin (Asia and South America). Red and orange colors indicate higher kinship and yellow colors indicate lower kinship

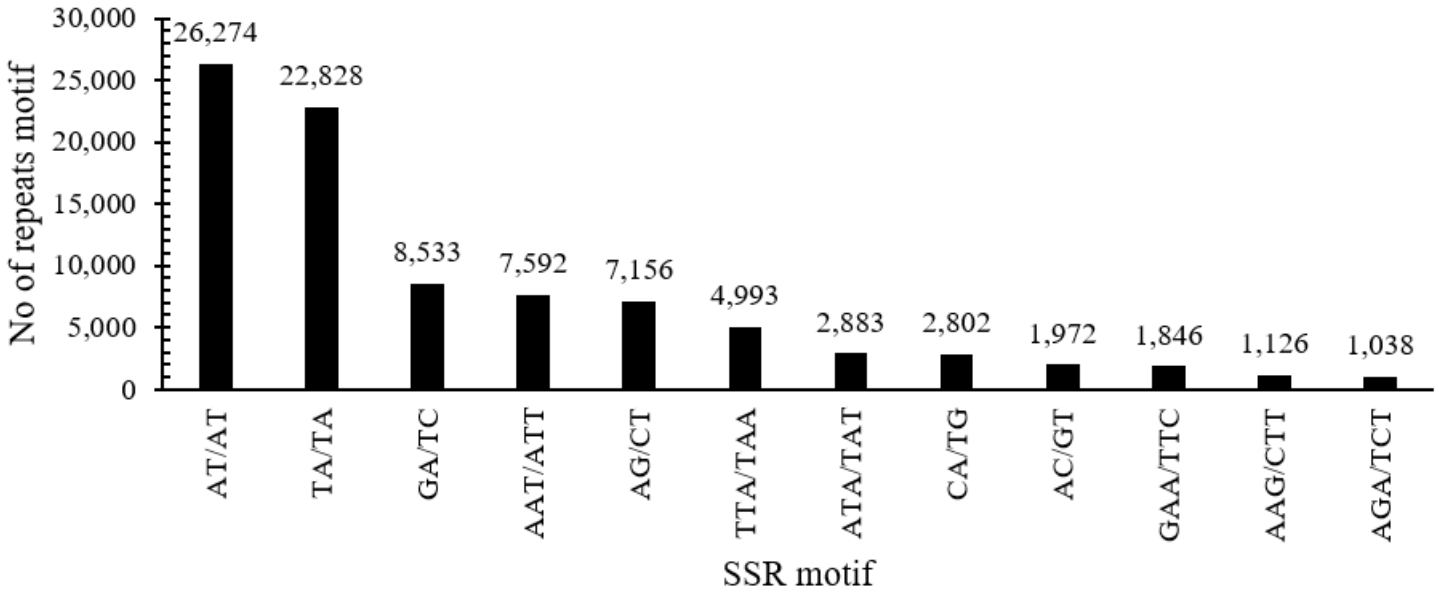

Figure 5

Frequency distribution of the six most frequent dinucleotide and trinucleotide SSR motifs. Only the six most frequent nucleotide motifs are shown.

Supplementary Files

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

- TableS1.docx