

Genome Wide Characterization and Comparative Analysis of Simple Sequence Repeats in Cucurbita Genomes

Lei Zhu

Henan Agricultural University <https://orcid.org/0000-0003-4210-1531>

Hua yu Zhu

Henan Agricultural University

Yan man Li

Henan Agricultural University

Xiang bin Wu

Henan Agricultural University

Jin tao Li

Henan Agricultural University

Zhen li Zhang

Henan Agricultural University

Yan jiao Wang

Henan Agricultural University

Jian bin Hu

Henan Agricultural University

Sen Yang

Henan Agricultural University

Yong qiang Li

Henan Agricultural University

Shou ru Sun (✉ ssr365@sina.com)

Lu Ming Yang

Henan Agricultural University

Research article

Keywords: Cucurbita species, SSR, the syntenic relationship, Genetic diversity, MAS

Posted Date: July 9th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-40696/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Genome wide characterization and comparative
analysis of simple sequence repeats in *Cucurbita*
genomes**

**Lei Zhu[†], Huayu Zhu[†], Yanman Li, Xiangbin Wu, Jintao Li, Zhenli Zhang,
Yanjiao Wang, Jianbin Hu, Sen Yang, Yongqiang Li, Shouru Sun^{*}, Luming
Yang^{*}**

College of Horticulture, Henan Agricultural University, 95 Wenhua Road,
Zhengzhou, 450002, China

[†] These authors contributed equally to this work

^{*} For correspondence email: the E-mail of the first corresponding author:

ssr365@sina.com, the E-mail of the second corresponding author:

luminyang@henau.edu.cn

Abstract

Background

The *Cucurbita* genus contains important economic crops in the world, while limited molecular markers have been developed in the past years. Simple sequence repeats (SSR) markers are powerful tools for the study of genetic mapping construction, genetic diversity analysis and genome wide association. The availability of pumpkin genome information has made it possible to analyze SSRs in genome wide across three *Cucurbita* species.

Results

In this paper, based on the whole genome sequences, 34,375 SSR loci were found in *C. moschata*, 30,577 SSR loci were found in *C. maxima* and 38,104 SSR loci were found in *C. pepo*. *C. pepo* has the maximum density of SSRs with an average of 145 SSR/Mb. In general, the frequency in total SSR loci decreased with the increase of the motif length, dinucleotide motifs were the most common motifs in the three species, and for the same repeat types, the SSR frequency decreased sharply with the increase of the repeat number. Most of those SSR loci were suitable for marker development (84.75% in *C. moscata*, 94.53% in *C. maxima* and 95.09% in *C. pepo*). Based on those markers, we compared and analyzed the cross-species SSR markers between *C. pepo* and other Cucurbitaceae species by silico-PCR. Using these cross-species primers, the high collinear relationships between *C. pepo* and the other two species were detected, respectively. Furthermore, the application of SSR markers in genetic diversity analysis was tested in *C. pepo*, the results showed that they were good tools to be used in genetic diversity analysis.

Conclusion

In this study, the genome wide SSR markers were detected from three *Cucurbita* species, and some of their applications were proved by comparative genomics and genetic diversity analysis. The large number of genome-wide SSR markers and cross-species markers would promote the basic and applied studies of *Cucurbita* species, such as gene mapping, QTLs mapping, comparative genomics and marker-assisted breeding.

Keywords: *Cucurbita* species, SSR, the syntenic relationship, Genetic diversity, MAS

Background

The *Cucurbita* genus ($2n=2x=40$), belonging to the Cucurbitaceae family, contains 13-30 species [1]. Most of these species are wild resources, and only three domesticated species including *Cucurbita. Maxima*, *Cucurbita. Moschata*, and *Cucurbita. Pepo* are widely cultivated in the world [2]. The earliest archaeological evidence for the domestication of *Cucurbita* species was about 8,000-10,000 years ago [3, 4]. The *Cucurbita* genus, native to the America and southern Mexico, was firstly introduced into European and then underwent a great diversification in Asia, which later became the secondary domestication center [3, 5, 6]. Nowadays, as an important economic crop, pumpkin has been widely grown all over the world. The largest cultivation area for pumpkin is Asia, followed by Europe and South America, and China is the main pumpkin production country in Asia. In 2012, the planting area of pumpkin in China was about 3.8×10^4 Hm², and the total output reached 7.0×10^6 tons (<http://faostat3.fao.org/>). *Cucurbita* species have a relatively smaller genome sizes and a greater chromosome numbers compared with other Cucurbitaceous crops [7], and the assembled genome sizes are 271.4 Mb for *C. maxima*, 269.9 Mb for *C. moschata*, and 263 Mb for *C. pepo*, respectively [8, 9]. With a long history of cultivation and domestication, the morphological variation of *Cucurbita* species shows a higher diversity in fruit shape, size and color than other Cucurbitaceae species [10]. Furthermore, *Cucurbita* species have strong roots and wide adaption to different adversities, such as cold stress, virus infection, and high salt, so they are widely used as rootstocks in grafting [11, 12].

Simple sequence repeats (SSR), because of their relative abundance, multi-allelism, co-dominance and cheap costs, are widely used in genetic mapping constructions, genetic diversity analysis and genome wide association studies [13, 14]. The whole genome sequencing of cucumber, melon and watermelon has been completed for years [15-17], and the genome-wide SSR markers in these crops have been characterized and developed, which have greatly promoted their application in gene and QTL mapping, comparative genomics [18-20]. For example, the syntenic relationships between melon and cucumber chromosomes were revealed using 199 SSR markers [21]. The genome syntenic relationships among watermelon, melon and

cucumber were confirmed by comparative mapping [22]. Based on massive genome-wide SSR markers, the complicated mosaic patterns of the chromosome synteny between melon and watermelon or cucumber were also established [23]. However, the developments of SSR markers in cucurbit species were limited. According to the conserved sequences among species and genera, a few of AFLP, RAPD and SSR markers were developed in previous studies [24-28]. These markers were still far from enough in the genetic diversity or comparative genomics researches. Although Esteras *et al.*, constructed the first *Cucurbita* genetic map based on the SNPs and found that the *Cucurbita* linkage groups were partially homoeologous to cucumber chromosomes using 304 SNPs and 11 SSR markers [29], the applications of these EST-SNP markers still had a great limitation. For example, the high cost of enzymes and the complicated operation procedure. Until recently, the related reports about *Cucurbita* genus gradually increased and the whole genome sequence of pumpkin was available. The whole-genome synteny analysis indicated that both *C. maxima* and *C. moschata* genomes underwent a whole genome duplication (WGD) event and pairs of *C. maxima* (or *C. moschata*) homoeologous regions shared between chromosomes corresponding to two sub-genomes [9]. Latterly, Montero *et al.* analyzed the genomic synteny between *C. pepo* and other Cucurbitaceae species (*Cucumis melo*, *Cucumis sativus* and *Citrullus lanatus*) and identified that the covered regions in most of the zucchini genome had a whole genome duplication again [8]. Furthermore, some transcriptomes of *Cucurbita* species were completed, and the EST-SSRs were developed [30-34]. The successive reports of *Cucurbita* genomic resources greatly promoted the large-scale development of SSR markers and gave us a benefit to construct the high-resolution maps and study the syntenic relationship and chromosomal rearrangements between *Cucurbita* species.

In this study, we carried out a genome wide identification of SSR motifs in three *Cucurbita* species, analyzed the distribution and frequency of different repeats types and studied the chromosome synteny of *Cucurbita pepo* with other Cucurbitaceae species. The derived SSR loci will be useful for population structure, genetic diversity, molecular assisted selection, map-based cloning and other studies in *Cucurbita* species.

Results

The frequency and distribution of different SSR types in *Cucurbita* genome

A total of 103,056 microsatellite sequences were identified in *Cucurbita* genome, including 34,375 SSR loci in the released 269.9 Mb genomic sequences of *C. moschata* cv. Rifu, 30,577 SSR loci in the released 271.4 Mb genomic sequences of *C. maxima* cv. Rimu and 38,104 SSR loci in the released 263 Mb genomic sequences of *C. pepo* MU-CU-16 (Table. 1). Obviously, *C. pepo* has the maximum density of SSRs with an average of 145 SSR/Mb.

Here, we analysed repeat types ranged from dinucleotide to octanucleotide. Among all of these nucleotide motifs, dinucleotide motifs (41.0%) seemed to be the most common motifs accounting for 41.78%, 39.90% and 41.01% of the total SSR loci discovered in the three genomes, respectively, followed by trinucleotide motifs (16.97%, 19.19%, and 17.88%, respectively), the octanucleotide motifs (3.78%, 3.76% , and 3.38%, respectively) were the least represented repeat types (Table. 1). In general, the frequency in total SSR loci decreased with increase of the motif length, except for heptanucleotide.

We further examined the SSR motif distribution with regard to their repeat numbers (Fig. 1). For all the repeat types, with increased of the repeat number, the SSR frequency decreased sharply, and this change was more obvious in the longer SSR motifs (Fig. 1). Consequently, the mean repeat numbers in dinucleotides were the highest in all the repeat types (Table. 1). The analysis of individual SSR types revealed that some of its own specific motifs were more prevalent than others in each class (Additional file 1, Fig S1). For example, The AT motif also was the most frequent motif in the three genomes, which accounting for 31.61% (in *C. moschata*), 28.81% (in *C. maxima*) and 30.45% (in *C. pepo*) of the total loci discovered. Similarly, the AAT, AAAT, AAAAT, AAAAAT, AAAAAAT, and AAAAAAAT motif (AATAATAT motif in *C. maxima*) were the most frequent types in each class, respectively (Additional file 1, Fig S1). These results revealed that AT-rich motifs were the most abundant ones in all SSR motifs in *C. moschata*, *C. maxima* and *C. pepo* genomes.

We investigated the SSR frequencies in each chromosome of the three *Cucurbit* species (Table. 2). The frequency of microsatellite loci was not correlated with the chromosome size. For example, in the *C. moschata* genome, the SSR density of the longest chromosome (Chr04) was just at the middle level, while Chr02 having a much shorter length than Chr04 had the highest SSR density. Similar things also happened in the other two genomes indicating that the distribution of SSRs was uneven in the Chromosomes (Table. 2). For better understanding of the distributions of different SSR motifs, we further checked their frequencies on each chromosome (Fig. 2). The results showed that different SSR types had the same tendency among different chromosomes, it seemed that the distribution of different SSR types on one chromosome was decided by their frequencies in the whole genome and the SSR density of the Chromosome (Fig. 2).

The genomics sequences containing these microsatellites were screened for PCR primer design, and 94,272 SSR microsatellite loci contained suitable flanking sites for SSR primer design. *C. moschata* had the lowest proportion of SSRs suitable for primers design (84.75%), while the percentages in *C. maxima* and *C. pepo* reached 94.53% and 95.09%, respectively (Table. 1). Though the dinucleotide repeat type was the most frequent one in all the three genomes, it didn't have a good performance in primers design. Hexanucleotide repeat type has the highest ratio of SSRs suitable for primers design, followed by pentanucleotide repeat types in all the three genomes, indicating that the longer motifs were more suitable for primers design in *Cucurbita* (Table. 1). Finally, 91,248 SSR primers (28,194 in *C. moschata*, 28,061 in *C. maxima* and 34,993 in *C. pepo*) were designed with some primers included more than one SSR loci as compound SSR (Additional file 2, Table S2-Table S4).

Chromosome synteny relationship of *Cucurbita pepo* with other Cucurbitaceae species

In order to understand the universality and correlation of SSR markers among Cucurbitaceae crops, we compared and analyzed the cross-species SSR markers between *C. pepo* and other Cucurbitaceae species by silico-PCR. We identified 391 cross-species SSR markers between *C. pepo* and *C. sativus*, 425 cross-species SSR markers between *C. pepo* and *C. melo*, 717 cross-species SSR markers between *C.*

pepo and *C. lanatus*, 11,732 cross-species SSR markers between *C. pepo* and *C. maxima*, and 15,274 cross-species SSR markers between *C. pepo* and *C. moschata* (Additional file 2, Table S5-S9). Collinear blocks to inversion-blocks ratio was 26:26 between *C. pepo* and *C. sativus* genomes, 25:36 between *C. pepo* and *C. melo* genomes, 51:38 between *C. pepo* and *C. lanatus* genomes, 154:158 between *C. pepo* and *C. maxima* genomes, and 153:152 between *C. pepo* and *C. moschata* genomes (Additional file 2, Table S5-S9). Interestingly, the ratio of collinear-blocks to inversion-blocks was nearly 1:1 among three *Cucurbita* species. Each *C. pepo* blocks shared 3-34 SSR markers with *C. sativus*, *C. lanatus* or *C. melo*. However, most of the *C. pepo* syntenic block shared larger number of SSR markers (3-1,080) with *C. maxima* or *C. moschata*. The *C. pepo* syntenic block CpeCmos53 had the largest number of shared SSR markers (1,080) between Cpe-Chr03 and Cmo-Chr14, which indicated that these regions were conservative during chromosome evolution.

By comparing the physical positions of those common shared-markers, the main syntenic relationships between *C. pepo* and other Cucurbitaceae species were listed in Table. 3, and the visualized chromosomes syntenic relationships of *C. pepo* with *C. lanatus*, *C. melo*, and *C. sativus* were shown in Fig. 3. The main syntenic relationships among chromosomes revealed complex mosaic patterns. In Fig. 3, each *C. pepo* chromosome was syntenic to more than two chromosomes in other Cucurbitaceae species. *C. pepo* chromosome Cpe9 and Cpe16 had the simplest syntenic pattern with watermelon, and each of them was mainly syntenic to one watermelon chromosome (Table. 3). Cpe9 was syntenic to watermelon chromosome W5, and 14 commonly shared SSR markers were found on them. From marker CpeSSR15544 to CpeSSR16107, there were three blocks and each block contained at least four SSR markers. According to the continuous physical positions of these markers on both of the reference genomes, the block CpeWM37 and CpeWM38 showed an inversion pattern and the block CpeWM39 showed a collinear pattern between *C. pepo* and *C. lanatus* (Additional file 2, Table S7). Similar comparison works were carried out between *C. pepo* and *C. sativus* or *C. pepo* and *C. melo* using the cross-species SSR markers. *C. pepo* chromosomes cpe7, cpe8, cpe11 and cpe20 had the simplest syntenic pattern with *C. sativus*, while each of them just syntenic to

one cucumber chromosome. Meanwhile, the simplest syntenic pattern between *C. pepo* and *C. melo* was mainly found on Cpe15, Cpe18, Cpe19 and Cpe20 (Table. 3). The most complicated syntenic pattern was found on Cpe1, which was corresponding to five chromosomes of *C. moschata*, four chromosomes of *C. maxima*, seven chromosomes of *C. lanatus*, three chromosomes of *C. sativus* and five chromosomes of *C. melon* (Table. 3).

Compared with the above analysis, the syntenic relationships between *C. pepo* and *C. moschata*, or between *C. pepo* and *C. maxima* were more simple and clear. Each of the 20 chromosomes in *C. pepo* was mainly syntenic with one chromosome in *C. moschata* and one chromosome in *C. maxima* (Table. 3). In order to see their syntenic relationships more directly, the single chromosome comparison analysis was carried out (Fig. 4). From Fig. 4, it was easy to find that each chromosome in *C. pepo* was collinear to one chromosome in *C. moschata* and one chromosome in *C. maxima*, these three chromosomes had highly consistency; the two chromosomes, individually from *C. moschata* and *C. maxima*, even were named as the same order in each genome. For example, Cpe1 was syntenic with Cmo4 and Cma4, Cpe2 was syntenic with Cmo1 and Cma1 (Fig. 4). Those results indicated that *C. pepo* had closer relationships with *C. moschata* and *C. maxima*, compared with other species in Cucurbitaceae; *C. moschata* and *C. maxima* might have a closer syntenic relationship than with *C. pepo*.

Application of SSR markers in *C. pepo* genetic diversity and population structure analysis

In a preliminary experiment, about four hundred SSR markers were screened using 61 materials. Based on the account of allelic number, the genomic coverage, and the PCR amplification efficiency, 66-core SSR markers were obtained (Additional file 2, Table S10). These markers exhibited clear band spectrums and distributed evenly on the chromosomes. In this study, 66 pairs of primers amplified 276 polymorphic sites in 61 zucchini materials with an average of 4.18 loci per pair of primers (Additional file 2, Table S11). The number of observed alleles (N_a) ranged from 2 to 9. The highest number of N_a (9) were observed by SSR010246, SSR026560, SSR026918, SSR027656, and SSR026980, followed by SSR011546, SSR003315 and SSR026797

with eight alleles. The number of effective alleles (N_e) varied from 1.03 to 6.07 with an average of 2.31. The Shannon information index (I) ranged from 0.083 to 1.96 with an average of 0.83. The PIC value ranged from 0.03 to 0.83 with an average of 0.43.

According to the results of the structure operation, it could be seen that when $K=2$, ΔK showed a significant peak value, indicating that 61 materials selected in this study could be divided into two groups (Additional file 1, Fig S2), named group I and group II. The group I contained 5 materials (8.20%), all of them were wild varieties; The group II contained 56 materials (91.80%), which were all cultivars (Fig. 5A). It indicated that our primers could clearly distinguish the cultivated materials from the wild varieties, and the backgrounds of the choosed materials were narrow. Furthermore, the phylogenetic tree was drawn by MEGA6.0 software (Fig. 5B), the 5 materials (2, 29, 30, 31 and 45) in group I were at the bottom of the phylogenetic tree, indicating that, based on the genetic analysis, the 61 zucchini materials were also divided into two clusters, which were completely in accordance with the result of structure analysis.

Discussion

Frequency, distribution and characterization of microsatellite in three *Cucurbita* genomes

With the development of the sequencing technology, the discovery and mining of genomic SSR loci had successful applications in many plant species such as cotton [35, 36], foxtail millet [37], cucumber [38], watermelon [39], tobacco [40] and melon [23]. Nevertheless, little effort has been paid on *Cucurbita* species. In this study, the density of the SSR markers in the genome of the *Cucurbita* species was about 113-145 SSR/Mb, *C. pepo* had the minimum length of the genome size and maximum number of microsatellites. The number of microsatellites and their density identified in this study was lower than that in cucumber (552 SSR/Mb), but was higher than that in melon (109 SSR/Mb) and watermelon (111 SSR/Mb) [23, 38, 39]. Except the natural difference among different genomes, there are many factors that can affect the deviations of the SSR density, different software, different parameter settings, and sequencing depths. For example, in this paper, basic motifs from dinucleotides to octanucleotides were checked with a minimum length of 18 bp (for di- and

tetranucleotides), 20 bp (for pentanucleotides), 24 bp (for hexanucleotides), 21 bp (for heptanucleotides), and 24 bp (for octanucleotides). In cucumber [38], though they also checked the motifs from dinucleotides to octanucleotides, their minimum lengths (12 bp for di- and tetranucleotides, 15 bp for pentanucleotides, 18 bp for hexanucleotides, 21 bp for heptanucleotides, and 24 bp for octanucleotides) were mostly shorter than our setting. The size of assembled grass genomes was higher than that in *Cucurbita* species, which leading to their higher SSR numbers than ours.

In this research, we analyzed the distribution and frequency of microsatellites in assembled *Cucurbita* genome. 36,234 polymorphic SSR markers in *C. pepo* identified by in silico analysis showed that 40% (14,489) and 18% (6,578) of them were dinucleotides and trinucleotides types, respectively (Table. 1). Previous studies have shown that the dinucleotides motifs with high repeat numbers were more polymorphic compared to those with short repeat units [41]. Consistent with the studies in watermelon and melon, the frequency analysis of various nucleotide repeat types in *Cucurbita* genus indicated that dinucleotide repeats were the most popular SSRs, followed by tri-, tetra-, penta-, hepta-, hexa- and octonucleotide repeats [23, 39]. However, this was different with the studies in other species. For example, the density of tetranucleotide repeats were the most abundant in *Cucumis sativus* (164.2 SSR/Mb), *Populus trichocarpa* (144.9 SSR/Mb), *Medicago truncatula* (102.8 SSR/Mb), and *Vitis vinifera* (171.3 SSR/Mb), and the trinucleotide repeats were the most abundant in *Arabidopsis thaliana* (146.6 SSR/Mb), *Glycine max* (103.1 SSR/Mb), and *Oryza sativa* (220.1 SSR/Mb) [38]. These results indicated that the most popular SSR motifs were varied in different species.

A general negative correlation was observed between the microsatellite frequency and the number of repeat units. Furthermore, AT or AAT type prevailed in the dicot plants [20]. Our results agreed with this result. Lately, SSR markers analysis in bitter gourd showed that all kinds of triple repeat units were the main type. A/T, AT/AT, AAT/ATT, and AAAT/ATTT were overrepresented and totally accounted for 63.86-76.30% of all motifs identified in the seven cucurbit genomes, respectively [42]. This pattern was also found in other genomes [18, 43, 44]. On the contrary, the frequency of the GC or CCG type was very low in the genomic level [45, 46]. GC, TC or GA

types have relatively stable structures. Most of the AT types distribute in non-genic regions and TC/GA types in coding sequences [40]. It has been reported that many bacteria SSRs in intergenic regions have regulating functions [47]. Whether these pumpkin SSR motifs played any role during specialization or protein coding was uncertain.

The chromosome synteny analysis between *C. pepo* and other Cucurbitaceae species by cross-species transferable markers

The chromosome synteny analysis has been carried out in many plants, such as cucumber, watermelon and melon [23], but little has been reported in *Cucurbita* genus. In this study, a large number of cross-species SSR markers were developed in *Cucurbita* genome, and they enabled us to investigate the syntenic blocks at a high resolution. Though the size of the pumpkin genome is similar to that of the other sequenced Cucurbitaceae species, the number of cross-species SSR markers in *Cucurbita* genus is much higher. Compared to hundreds of shared markers in previous studies [21], much more cross-species transferable SSR markers in *Cucurbita* genus were used for chromosome synteny analysis in our study. It might be the case that there had a WGD event leading to the high abundance of SSR markers [8, 9], but it's not observed in other sequenced Cucurbitaceae species, such as cucumber [15], melon [48], watermelon [16].

After the genome comparison between *C. pepo* and cucumber, *C. pepo* and melon, *C. pepo* and watermelon, 52, 61 and 89 syntenic blocks were identified, respectively. These blocks distributed on all chromosomes. The similarly works among *Cucurbita* species was also reported in previous study, about 63.2%, 58.7%, and 68.3% of the *C. maxima* genomic regions were syntenic to melon, cucumber, and watermelon, respectively, while those numbers changed to 64.0%, 62.2%, and 69.5% in *C. moschata*, and they were composed of homoeologous blocks [9]. In most cases, because of the chromosome fission, the multiple synteny relationships between *C. pepo* and other Cucurbitaceae species were shown. The most complicated syntenic pattern existed in Chr1 of *C. pepo*, which was syntenic to seven watermelon chromosomes, indicating that complex structure changes occurred after their divergence from the ancestor. The ratio of collinear-blocks to inversion-blocks was

nearly 1:1 in *Cucurbita* genus, the possible reason was that the genome duplication and interchromosomal exchanges happened randomly [9].

According to the cross-species SSR markers analysis, we identified more syntenic blocks in *Cucurbita* species than that in melon, cucumber or watermelon. At the same time, we found that each block contained abundant SSR markers. For example, the *C. pepo* syntenic block CpeCmos53 had the largest number of shared SSR markers (1,080). As we all known, the largest number of shared SSR markers (386) between melon and cucumber was found in block C3 [23]. This indicated that there were lots of genetic variations and highly evolutionary relationships among homologous species. The single chromosomal analysis by cross-species shared SSR markers showed that there were six large-scale inversion regions in different chromosomes between *C. pepo* and *C. moschata* or between *C. pepo* and *C. maxima*. The previous study reported that all chromosomes except Chr4 contained only homologous blocks between *C. maxima* and *C. moschata* [9]. This might be due to genome duplication, large-scale inter-chromosomal exchanges or long-term evolutionary forces. It also indicated that *C. pepo* had more complex evolutionary processes than other *Cucurbita* species.

The genetic diversity and population structure of *C. pepo*

As one of the species with the most complex genetic background, limited molecular markers have been used for genetic diversity and population structure analysis in *C. pepo*. For example, *Ferriol et al.* used SRAP and AFLP markers to analyze the population structure in 69 *C. pepo* species, and found that SRAP markers were more consistent with the morphological character [49]. *Nonthuko et al.* used nine RAPD markers and ten SSR primers to study genetic diversity in seven *C. pepo* inbred lines and detected 100 and 56 alleles, respectively, but only the RAPD markers could distinguish these cultivars according to their fruit colors [50]. *Paris et al.* used AFLP, ISSR and SSR markers to analyze the relationship between wild and domestic species, and successfully clustered them into three groups based on their fruit colors, fruit sizes and origins [26]. Because of the scarcity of highly polymorphic markers and genome-sequencing data, most previous studies in *Cucurbita* species have been done using low throughput and anonymous markers. In this study, we developed 91,248

SSR markers from the three *Cucurbita* genomes, they would benefit the genetic studies in the future.

Genome wide SSR markers were not developed until the completion of the genomic sequencing in *Cucurbita* species. In this study, we employed 66 pairs of high polymorphic SSR markers to investigate the genetic diversity of 61 *C. pepo* materials and an average of 4.18 alleles were amplified. The UPGMA cluster analysis results were almost consistent with the results of population structure analysis (Fig. 5). In this study, we have successfully clustered these materials into two subspecies, *subsp. Ovifer* (or *subsp.texana*) and *subsp. pepo*. Most of *C. pepo subsp. pepo* accessions were clustered into group II, and the five *C. pepo subsp. Ovifer* accessions were clustered into group I. This result was consistent with the previous studies [29, 49], but Decker classified *C. pepo* into three subspecies, *C. pepo subsp. fraterna* (Bailey) Andres, *C. pepo subsp.texana* (Scheele) Filov and *C. pepo subsp. Pepo* [51]. *C. pepo subsp. fraterna* has been considered as the putative ancestor for the species as a whole and consists exclusively of wild gourds from northeastern Mexico [52]. Moreover, the genetic structure revealed that the background of some materials was a mixed model between group I and group II. For example, the material 45 in group I has complex genetic background, just like the material 14 and 16 in group II, indicating that these accessions may have undergone gene exchange between two subspecies.

The *Cucurbita* species is an economically important crop, but its breeding process fall behind the other Cucurbitaceous species. Limited high-quality cultivars cannot meet the production requirements. Thus, speeding the current breeding programs using MAS become more and more important. The whole-genome SSR markers detected in this study will promote the development and the utilization in basic and applied research.

Conclusions

In this paper, based on reference sequences of three *Cucurbita* species, a total of 91,248 SSR markers were developed at the whole genome level. Then their frequency and distribution were detected and analysed. According to these markers, the cross-

species SSR markers were checked by silico-PCR and the synteny relationships were analysed between *C.pepo* and other Cucurbitaceae species. Furthermore, 66 polymorphism SSR markers were employed to check the genetic diversity of *C.pepo*. These SSR markers will benefit the related researches and accelerate the process of pumpkin breeding.

Methods

Plant Materials

All the pumpkin materials used in this study were introduced from the national crop germplasm resource platform of china (sub-platform of vegetable germplasm resource) in 2018, among which 4 materials came from Russia, 1 from America and 56 materials came from 17 provinces of China. The number and sources were shown in Table S1.

Genome SSR identification and development in Cucurbitaceae

The genome information of watermelon, melon, cucumber and pumpkin was downloaded from <http://cucurbitgenomics.org/>. To develop a set of higher polymorphism SSR primers for the future study, the criteria used for microsatellite identification in this study was from 2 to 8 bp, and mononucleotides were not considered due to the difficulty in distinguishing bona fide microsatellites from sequencing or assembly error. The Microsatellite Identification tool (MISA) was used to identify and analyze SSR markers, including perfect and compound microsatellites. Specific screening details were as follows: repeats with a minimum length of 18 bp (for di- and tetranucleotides), 20 bp (for pentanucleotides), 24 bp (for hexanucleotides), 21 bp (for heptanucleotides), and 24 bp (for octanucleotides). The oligonucleotide primers for these SSRs were designed according to the flanking genomic sequence using Primer3 software (v.1.1.4). Primers were designed to generate amplicons of 100-300 bp in length with the following minimum, optimum and maximum values for Primer3 parameters: primer length (bp): 18-20-24; Tm (°C): 50-55-60. Other parameters used the default program values.

In silico PCR and synteny analysis of SSR markers in Cucurbitaceae

Using the zucchini (*C. pepo* MU-CU-16) genome SSR markers as a reference, we comparatively analyzed the genome SSR information of cucumber (Gy14), melon (DH92), watermelon (97103), *C. moschata* cv. Rifu and *C. maxima* cv. Rimu, respectively. This was performed with a custom Perl script that used the NCBI BLASTN program as a search engine with expect value of 10 and filtering. We allowed up to five nucleotide mismatches at the 5' end of the primer but no mismatches at the 3' end, and a minimum of 90% overall match homology. To establish the syntenic relationships of chromosomes between *C. pepo* and *C. sativus*, *C. lanatus*, *C. melo*, *C. maxima* or *C. moschata*. We discarded the repeat SSR markers in different genomes, only kept the SSR markers in these genomes, which had single silico-PCR product. Besides, these shared SSR markers located on the chromosomal unanchored scaffolds were further filtered. The SSR marker-based syntenic relationships were finally visualized with visualization blocks in Circos software v-0.55 [53].

Genomic DNA extraction, PCR amplification and electrophoresis detection

Genomic DNA of all the materials was extracted using 1 g young leaf sample with the CTAB method. The extracted DNA was dissolved in 1×TE buffer. The concentration and purity were detected by Nanodrop-2000 nucleic acid analyzer. The extracted DNA was diluted to 30 ng/μl as a working solution, and kept at 4 °C.

Each PCR reaction contained 1 μl of template DNA, 0.5 μl each of forward and reverse primers, 5 μl mastermix and 3 μl ddH₂O. The amplification was carried out as follows: an initial denaturing step at 95 °C for 5 min, 94 °C for 30 s, followed by 6 cycles of 68-58 °C for 45 s, each cycle was reduced by 2 °C, each annealing time was 1 min, and 72 °C for 1 min; 30 cycles of 94 °C for 30 s, 50 °C for 30 s, and 72 °C for 1 min; In the last cycle, primer extension was performed at 72 °C for 10 min.

PCR products were analyzed by 9% polyacrylamide gel electrophoresis, and 100 bp DNA ladder was used as the reference marker. After electrophoresis, silver staining was performed to display the PCR products, and photos were taken for preservation.

Cluster calculating

The heterozygosity (H_e), observed gene number (N_a), effective alleles (N_e), observed heterozygosity (H_o) and Shannon-Weaver index (I) were calculated by the pop-gen

software. Polymorphism information content (PIC) of SSR markers was computed by EXCEL. When the PIC of a SSR marker was below 0.25, it was considered as a low polymorphism marker, and it was considered as a high polymorphism marker if its PIC was above 0.5.

These amplification bands of each SSR-PCR primer pairs were separated by polyacrylamide gel electrophoresis. The number of alleles stood by the total number of these bands in this SSR loci. In the same location, having band was marked as "1", no band was marked as "0", and the missing band was marked as "-1". In this paper, we used Genalex-6 software to conduct the matrix calculation of SSR marker data which had been assigned a value, then transformed it into triangle matrix, saved it as a mega-file, finally, imported the mega-file into the Mega-6.0 software, and selected UPGMA algorithm in "phylogeny" drop-down menu to draw the cluster diagram [54].

The software Structure V2.3 was used to analyze the population structure [55, 56]. An admixture model and correlated allele frequencies was used to estimate the number of the populations. For each of the K-values (range from 1 to 5), ten independent runs were performed with a burn-in period of 100,000 followed by 500,000 Markov Chain Monte Carlo. The optimal K-values depends on the peak of $K = \text{mean}(|\ln P(D)|) / (\text{sd} \ln P(D))$. Based on the structure results, the most probable K-value was analysis by the Structure Harvester (http://taylor0.biology.ucla.edu/struct_harvest/).

List of abbreviations

SSR, simple sequence repeats; QTL, quantitative trait loci; AFLP, amplified fragment length polymorphism; RAPD, random amplified polymorphic DNA; SNP, single nucleotide polymorphism; WGD, whole genome duplication; PCR, polymerase chain reaction; ISSR, Inter-simple sequence repeat; UPGMA, unweighted Pair Group Method with Arithmetic Mean.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data generated or analysed during this study are reflected in the present article. Our datasets of materials and genome-wide SSR markers were listed in Additional file 2.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by Science and Technology Project of Henan Province (agriculture) (182102110280) and Innovation and Entrepreneurship Training Program for College Students (201910466017), China.

Authors' contributions

LZ and HYZ performed the data analysis and wrote the manuscript. YML and XBW conducted validation of SSR polymorphism, genetic diversity and population structure analysis. JTL, ZLZ and YJW participated in SSR genetic diversity analysis. JBH, SY and YQL involved in the germplasm collection and growth. SSR and LMY designed this paper and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors are highly grateful to Professor Yufeng Wu of Nanjing Agricultural University in Circos analysis. The *C.pepo* materials used in this study were introduced from the national crop germplasm resource platform of china. Our thanks for the help of Jiangping Song from Chinese Academy of Agricultural Sciences in material's

introduction. We are also indebted to the editor and reviewers for critically evaluating the manuscript and providing constructive comments for its improvement.

Reference:

1. **Toxic plants of North America.** *Choice: Current Reviews for Academic Libraries* 2013, **50**(12):2202-2203.
2. Loy JB: **Morpho-physiological aspects of productivity and quality in squash and pumpkins (*Cucurbita spp.*).** *Crit Rev Plant Sci* 2004, **23**(4):337-363.
3. Nee M: **The domestication of *cucurbita* (Cucurbitaceae).** *Econ Bot* 1990, **44**(3):56-68.
4. Smith BD: **The Initial Domestication of *Cucurbita pepo* in the Americas 10,000 Years Ago.** *Science* 1997, **276**(5314):932-934.
5. Ferriol M, Picó B: **Pumpkin and Winter Squash.** In: *Vegetables I: Asteraceae, Brassicaceae, Chenopodiaceae, and Cucurbitaceae.* Edited by Prohens J, Nuez F. New York, NY: Springer New York; 2008: 317-349.
6. Kates HR, Soltis PS, Soltis DE: **Evolutionary and domestication history of *Cucurbita* (pumpkin and squash) species inferred from 44 nuclear loci.** *Mol Phylogenet Evol* 2017, **111**:98-109.
7. Arumuganathan K, Earle ED: **Nuclear DNA content of some important plant species.** *Plant Molecular Biology Reporter* 1991, **9**(3):208-218.
8. Montero-Pau J, Blanca J, Bombarely A, Ziarsolo P, Esteras C, Marti-Gomez C, Ferriol M, Gomez P, Jamilena M, Mueller L *et al*: **De novo assembly of the zucchini genome reveals a whole-genome duplication associated with the origin of the *Cucurbita* genus.** *Plant Biotechnol J* 2018, **16**(6):1161-1171.
9. Sun HH, Wu S, Zhang GY, Jiao C, Guo SG, Ren Y, Zhang J, Zhang HY, Gong GY, Jia ZC *et al*: **Karyotype Stability and Unbiased Fractionation in the Paleo-Allotetraploid *Cucurbita* Genomes.** *Mol Plant* 2017, **10**(10):1293-1306.
10. Savage JA, Haines DF, Holbrook NM: **The making of giant pumpkins: how selective breeding changed the phloem of *Cucurbita maxima* from source to sink.** *Plant Cell Environ* 2015, **38**(8):1543-1554.
11. Davis AR, Perkins-Veazie P, Sakata Y, Lopez-Galarza S, Maroto JV, Lee SG, Huh YC, Sun ZY, Miguel A, King SR *et al*: **Cucurbit grafting.** *Crit Rev Plant Sci* 2008, **27**(1):50-74.
12. Lee JM, Kubota C, Tsao SJ, Bie Z, Echevarria PH, Morra L, Oda M: **Current status of vegetable grafting: Diffusion, grafting techniques, automation.** *Sci Hortic-Amsterdam* 2010, **127**(2):93-105.
13. Lv J, Qi JJ, Shi QX, Shen D, Zhang SP, Shao GJ, Li H, Sun ZY, Weng YQ, Shang Y *et al*: **Genetic Diversity and Population Structure of Cucumber (*Cucumis sativus* L.).** *Plos One* 2012, **7**(10).

14. Li Y, Wen C, Weng Y: **Fine mapping of the pleiotropic locus B for black spine and orange mature fruit color in cucumber identifies a 50 kb region containing a R2R3-MYB transcription factor.** *Theor Appl Genet* 2013, **126**(8):2187-2196.
15. Huang SW, Li RQ, Zhang ZH, Li L, Gu XF, Fan W, Lucas WJ, Wang XW, Xie BY, Ni PX *et al*: **The genome of the cucumber, *Cucumis sativus* L.** *Nat Genet* 2009, **41**(12):1275-U1229.
16. Guo SG, Zhang JG, Sun HH, Salse J, Lucas WJ, Zhang HY, Zheng Y, Mao LY, Ren Y, Wang ZW *et al*: **The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions.** *Nat Genet* 2013, **45**(1):51-+.
17. Jordi Garcia-Mas AB, Walter Sanseverino, Michael Bourgeois, Gisela Mir, Víctor M. González, Elizabeth Hénaff, Francisco Câmara, Luca Cozzuto, Ernesto Lowy, Tyler Alioto, Salvador Capella-Gutiérrez, Jose Blanca, Joaquín Cañizares, Pello Ziarsolo, Daniel Gonzalez-Ibeas, Luis Rodríguez-Moreno, Marcus Droege, Lei Du, Miguel Alvarez-Tejado, Belen Lorente-Galdos, Marta Melé, Luming Yang, Yiqun Weng, Arcadi Navarro, Tomas Marques-Bonet, Miguel A. Aranda, Fernando Nuez, Belén Picó, Toni Gabaldón, Guglielmo Roma, Roderic Guigó, Josep M. Casacuberta, Pere Arús, and Pere Puigdomènech **The genome of melon (*Cucumis melo* L.).** *Proceedings of the National Academy of Sciences* 2012.
18. Cavagnaro PF, Senalik DA, Yang L, Simon PW, Harkins TT, Kodira CD, Huang S, Weng Y: **Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.).** *BMC Genomics* 2010, **11**:569.
19. Zhu H, Guo L, Song P, Luan F, Hu J, Sun X, Yang L: **Development of genome-wide SSR markers in melon with their cross-species transferability analysis and utilization in genetic diversity study.** *Mol Breeding* 2016, **36**(11):153.
20. Zhu H, Song P, Koo DH, Guo L, Li Y, Sun S, Weng Y, Yang L: **Genome wide characterization of simple sequence repeats in watermelon genome and their application in comparative mapping and genetic diversity analysis.** *BMC Genomics* 2016, **17**:557.
21. Li DW, Cuevas HE, Yang LM, Li YH, Garcia-Mas J, Zalapa J, Staub JE, Luan FS, Reddy U, He XM *et al*: **Syntenic relationships between cucumber (*Cucumis sativus* L.) and melon (*C. melo* L.) chromosomes as revealed by comparative genetic mapping.** *Bmc Genomics* 2011, **12**.
22. Yang LM, Koo DH, Li DW, Zhang T, Jiang JM, Luan FS, Renner SS, Henaff E, Sanseverino W, Garcia-Mas J *et al*: **Next-generation sequencing, FISH mapping and synteny-based modeling reveal mechanisms of decreasing dysploidy in Cucumis.** *Plant J* 2014, **77**(1):16-30.
23. Zhu HY, Guo LQ, Song PY, Luan FS, Hu JB, Sun XF, Yang LM: **Development of genome-wide SSR markers in melon with their cross-species transferability analysis and utilization in genetic diversity study.** *Mol Breeding* 2016, **36**(11).
24. Hi LY, Jeong JH, Hong KH, Dong KB: **Use of Random Amplified Polymorphic DNAs for Linkage Group Analysis in Interspecific Hybrid F₂ Generation of *Cucurbita*.** *Horticulture Environment & Biotechnology* 1995, **36**(3):323-330.

25. Brown RN, Myers JR: **A genetic map of squash (*Cucurbita* sp.) with randomly amplified polymorphic DNA markers and morphological markers.** *Journal of the American Society for Horticultural Science* 2002, **127**(4):568-575.
26. Paris HS, Yonash N, Portnoy V, Mozes-Daube N, Tzuri G, Katzir N: **Assessment of genetic relationships in *Cucurbita pepo* (Cucurbitaceae) using DNA markers.** *Theor Appl Genet* 2003, **106**(6):971-978.
27. Zraidi A, Stift G, Pachner M, Shojaeiyan A, Gong L, Lelley T: **A consensus map for *Cucurbita pepo*.** *Mol Breeding* 2007, **20**(4):375-388.
28. Gong L, Stift G, Kofler R, Pachner M, Lelley T: **Microsatellites for the genus *Cucurbita* and an SSR-based genetic linkage map of *Cucurbita pepo* L.** *Theoretical and Applied Genetics* 2008, **117**(1):37-48.
29. Esteras C, Gomez P, Monforte AJ, Blanca J, Vicente-Dolera N, Roig C, Nuez F, Pico B: **High-throughput SNP genotyping in *Cucurbita pepo* for map construction and quantitative trait loci mapping.** *Bmc Genomics* 2012, **13**.
30. Blanca J, Canizares J, Roig C, Ziarsolo P, Nuez F, Pico B: **Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (Cucurbitaceae).** *Bmc Genomics* 2011, **12**.
31. Wyatt LE, Strickler SR, Mueller LA, Mazourek M: **An acorn squash (*Cucurbita pepo* ssp. *ovifera*) fruit and seed transcriptome as a resource for the study of fruit traits in *Cucurbita*.** *Hortic Res-England* 2015, **2**.
32. Xanthopoulou A, Psomopoulos F, Ganopoulos I, Manioudaki M, Tsaftaris A, Nianiou-Obeidat I, Madesis P: **De novo transcriptome assembly of two contrasting pumpkin cultivars.** *Genom Data* 2016, **7**:200-201.
33. Aliko X, Ganopoulos I, Psomopoulos F, Manioudaki M, Moysiadis T, Kapazoglou A, Osathanunkul M, Michailidou S, Kalivas A, Tsaftaris A *et al*: **De novo comparative transcriptome analysis of genes involved in fruit morphology of pumpkin cultivars with extreme size difference and development of EST-SSR markers,** vol. 622; 2017.
34. Vitiello Alessia SD, D'Agostino Nunzio, Digilio Maria Cristina, Pennacchio, Francesco CG, Rao Rosa: **Unraveling zucchini transcriptome response to aphids.** *Peer J PrePrints* 2016, **4**.
35. Lu CR, Zou CS, Zhang YP, Yu DQ, Cheng HL, Jiang PF, Yang WC, Wang QL, Feng XX, Prosper MA *et al*: **Development of chromosome-specific markers with high polymorphism for allotetraploid cotton based on genome-wide characterization of simple sequence repeats in diploid cottons (*Gossypium arboreum* L. and *Gossypium raimondii* Ulbrich).** *Bmc Genomics* 2015, **16**.
36. Wang Q, Fang L, Chen JD, Hu Y, Si ZF, Wang S, Chang LJ, Guo WZ, Zhang TZ: **Genome-Wide Mining, Characterization, and Development of Microsatellite Markers in *Gossypium* Species.** *Sci Rep-Uk* 2015, **5**.
37. Zhang S, Tang CJ, Zhao Q, Li J, Yang LF, Qie LF, Fan XK, Li L, Zhang N, Zhao MC *et al*: **Development of highly polymorphic simple sequence repeat markers using genome-wide microsatellite variant analysis in Foxtail millet [*Setaria italica* (L.) P. Beauv.].** *Bmc Genomics* 2014, **15**.

38. Cavagnaro PF, Senalik DA, Yang LM, Simon PW, Harkins TT, Kodira CD, Huang SW, Weng YQ: **Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.).** *Bmc Genomics* 2010, **11**.
39. Zhu HY, Song PY, Koo DH, Guo LQ, Li YM, Sun SR, Weng YQ, Yang LM: **Genome wide characterization of simple sequence repeats in watermelon genome and their application in comparative mapping and genetic diversity analysis.** *Bmc Genomics* 2016, **17**.
40. Wang XW, Yang S, Chen YD, Zhang SM, Zhao QS, Li M, Gao YL, Yang L, Bennetzen JL: **Comparative genome-wide characterization leading to simple sequence repeat marker development for *Nicotiana*.** *Bmc Genomics* 2018, **19**.
41. Weber JL: **Informativeness of human (dC-dA)n.(dG-dT)n polymorphisms.** *Genomics* 1990, **7(4):524-530**.
42. Cui J, Cheng J, Nong D, Peng J, Hu Y, He W, Zhou Q, Dhillon NPS, Hu K: **Genome-Wide Analysis of Simple Sequence Repeats in Bitter Melon (*Momordica charantia*).** *Front Plant Sci* 2017, **8**.
43. Kim TS, Booth JG, Gauch HG, Jr., Sun Q, Park J, Lee YH, Lee K: **Simple sequence repeats in *Neurospora crassa*: distribution, polymorphism and evolutionary inference.** *Bmc Genomics* 2008, **9:31**.
44. Cheng J, Zhao Z, Li B, Qin C, Wu Z, Trejo-Saavedra DL, Luo X, Cui J, Rivera-Bustamante RF, Li S *et al*: **A comprehensive characterization of simple sequence repeats in pepper genomes provides valuable resources for marker development in *Capsicum*.** *Sci Rep* 2016, **6:18919**.
45. Tangphatsornruang S, Somta P, Uthapaisanwong P, Chanprasert J, Sangsrakru D, Seehalak W, Sommanas W, Tragoonrun S, Srinives P: **Characterization of microsatellites and gene contents from genome shotgun sequences of mungbean (*Vigna radiata* (L.) Wilczek).** *Bmc Plant Biol* 2009, **9:137**.
46. Portis E, Lanteri S, Barchi L, Portis F, Valente L, Toppino L, Rotino GL, Acquadro A: **Comprehensive Characterization of Simple Sequence Repeats in Eggplant (*Solanum melongena* L.) Genome and Construction of a Web Resource.** *Front Plant Sci* 2018, **9**.
47. Zhao ZX, Guo C, Sutharzan S, Li P, Echt CS, Zhang J, Liang C: **Genome-Wide Analysis of Tandem Repeats in Plants and Green Algae.** *G3-Genes Genom Genet* 2014, **4(1):67-78**.
48. Garcia-Mas J, Benjak A, Sanseverino W, Bourgeois M, Mir G, Gonzalez VM, Henaff E, Camara F, Cozzuto L, Lowy E *et al*: **The genome of melon (*Cucumis melo* L.).** *P Natl Acad Sci USA* 2012, **109(29):11872-11877**.
49. Ferriol M, Pico B, Nuez F: **Genetic diversity of a germplasm collection of *Cucurbita pepo* using SRAP and AFLP markers.** *Theoretical And Applied Genetics* 2003, **107(2):271-282**.
50. Ntuli NR, Tongoona PB, Zobolo AM: **Genetic diversity in *Cucurbita pepo* landraces revealed by RAPD and SSR markers.** *Sci Hortic-Amsterdam* 2015, **189:192-200**.
51. Decker DS: **Origin (s), Evolution, and Systematics of *Cucurbita pepo* (Cucurbitaceae).** *Econ Bot* 1988, **42:4-15**.

52. Paris HS: **Germplasm enhancement of *Cucurbita pepo* (pumpkin, squash, gourd: Cucurbitaceae): progress and challenges.** *Euphytica* 2015, **208**(3):415-438.
53. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: **Circos: an information aesthetic for comparative genomics.** *Genome Res* 2009, **19**(9):1639-1645.
54. Peakall R, Smouse PE: **GenALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update.** *Bioinformatics* 2012, **28**(19):2537-2539.
55. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data.** *Genetics* 2000, **155**(2):945-959.
56. Evanno G, Regnaut S, Goudet J: **Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study.** *Mol Ecol* 2005, **14**(8):2611-2620.

Figures

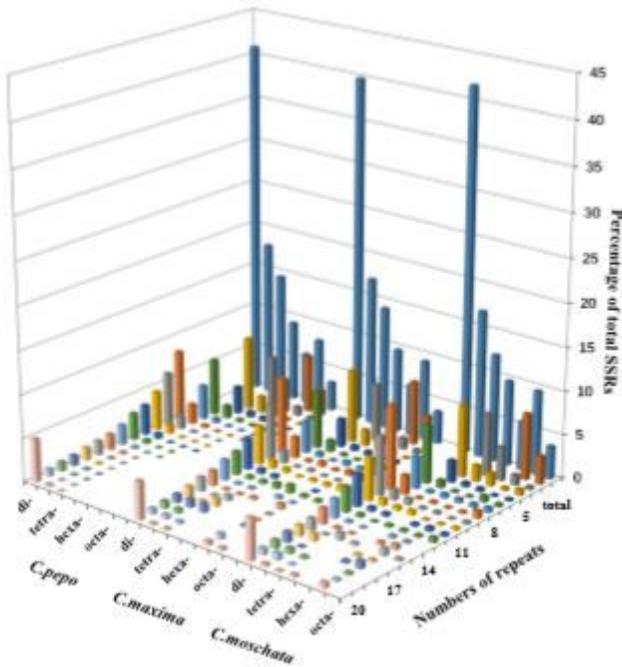


Figure 1

Distribution of SSR motif repeat numbers and relative frequency in Cucurbita genome. The vertical axis shows the abundance of microsatellites that have different motif repeat numbers (from 3 to > 20) with different colors.

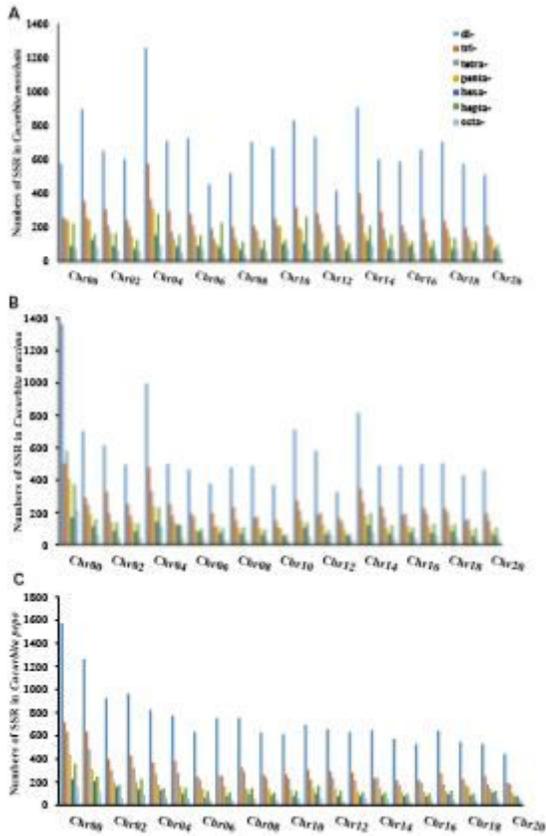


Figure 2

The distribution of SSR repeat types on each chromosome in *C.moschata*. The vertical axis shows the number of microsatellites from di-nucleotide to octo-nucleotide which are discriminated by different colors. The horizontal axis show different chromosomes of *C.moschata* and chr0 means all the chromosome unanchored scaffolds. B, The distribution of SSR repeat types on each chromosome in *C.maxima*. The vertical axis shows the number of microsatellites from di-nucleotide to octo-nucleotide which are discriminated by different colors. The horizontal axis show different chromosomes of *C.maxima* and chr0 means all the chromosome unanchored scaffolds. C, The distribution of SSR repeat types on each chromosome in *C.pepo*. The vertical axis shows the number of microsatellites from di-nucleotide to octo-nucleotide which are discriminated by different colors. The horizontal axis show different chromosomes of *C.pepo* and LG00 means all the chromosome unanchored scaffolds.

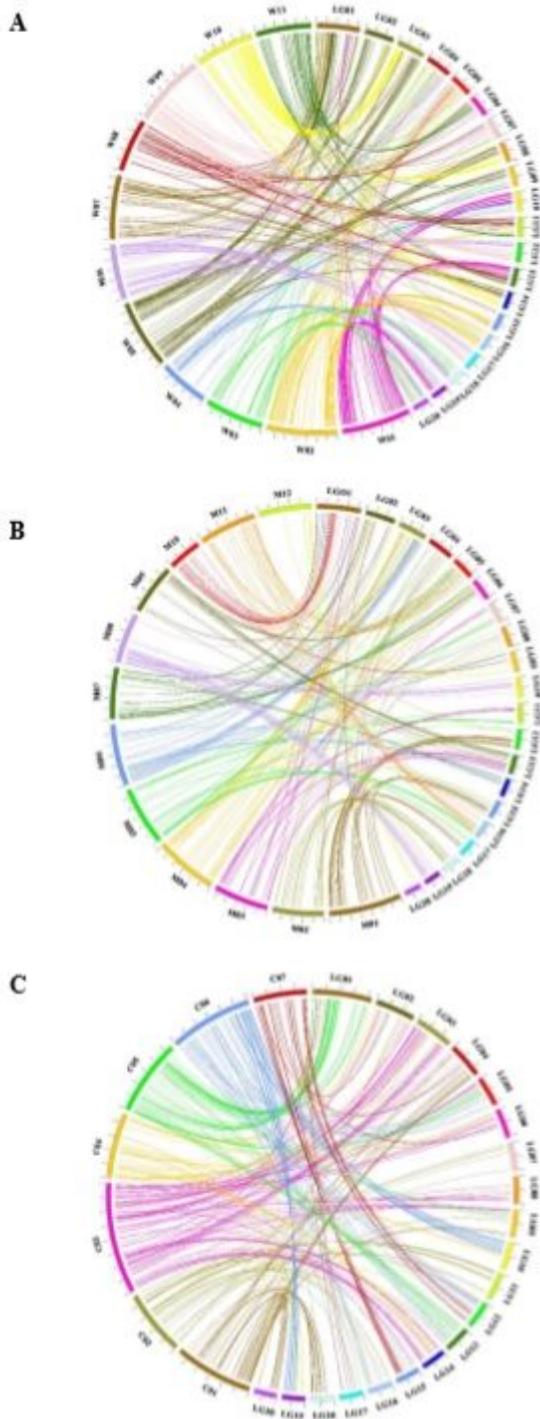


Figure 3

Syntenic relationships of *C. pepo* with *C. lanatus* (A), *C. melo* (B) and *C. sativus* (C). Chromosome synteny between *C. pepo* and *C. sativus* is based on 391 cross-species markers; synteny between *C. pepo* and *C. melo* is based on 425 cross-species markers; synteny between *C. pepo* and *C. lanatus* is based on 717 cross-species markers. W1-W11 represent *C. lanatus* eleven chromosomes, M01-M12 represent *C. melo* twelve chromosomes, C01-C07 represent *C. sativus* seven chromosomes, and LG01-LG20 represent *C. pepo*

twenty chromosomes. Syntenic blocks are connected by with the same color lines from *C.pepo* chromosomes.

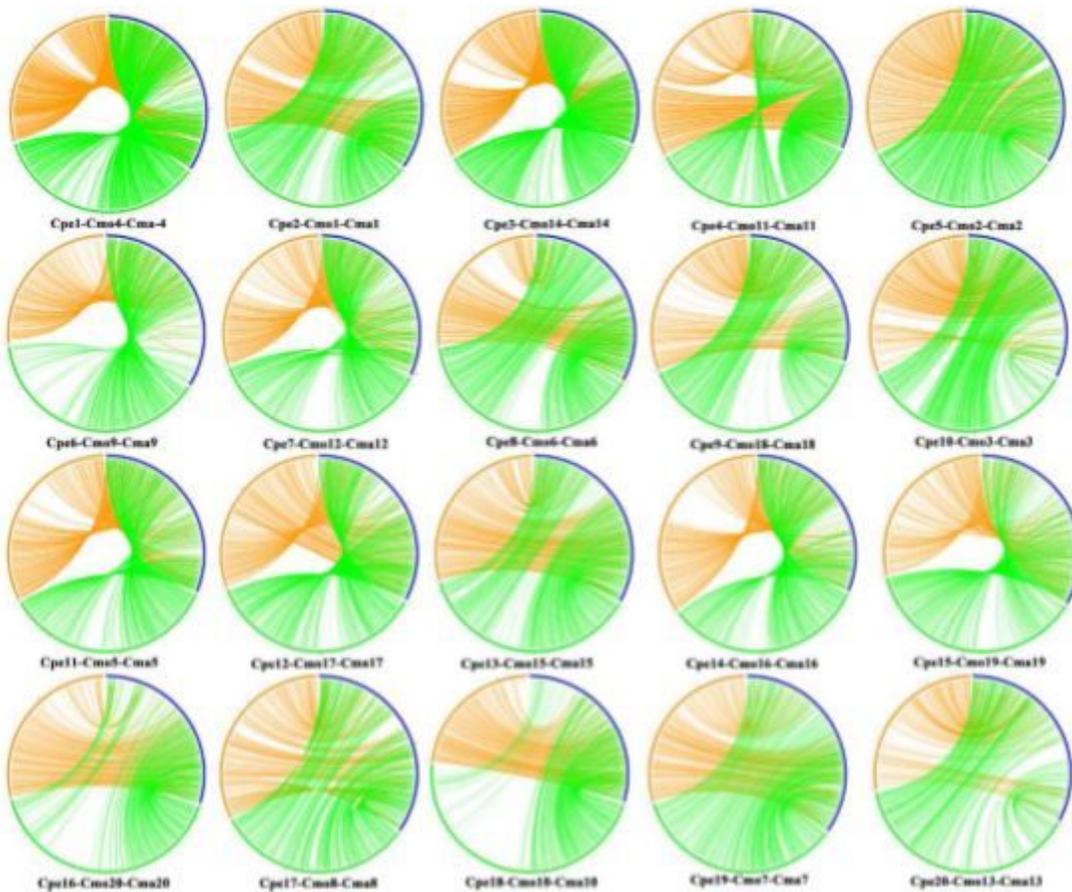


Figure 4

Chromosome synteny of *C.pepo* (blue) with *C.moschata* (green) and *C.maxima* (yellow). The physical positions of chromosomes of each crop in the figure are arranged clockwise. Chromosome synteny between *C.pepo* and *C.moschata* is based on 14276 cross-species markers; synteny between *C.pepo* and *C.maxima* is based on 10655 cross-species markers. Cpe1-Cpe20 represent *C.pepo*-twenty chromosomes, Cmo1-Cmo20 represent *C.moschata* chromosomes and Cma1-Cma20 represent *C.maxima* chromosomes. The syntenic relationship between *C.pepo* and *C.moschata* are connected with the green color lines, and the syntenic relationship between *C.pepo* and *C.maxima* are connected with the yellow color lines.

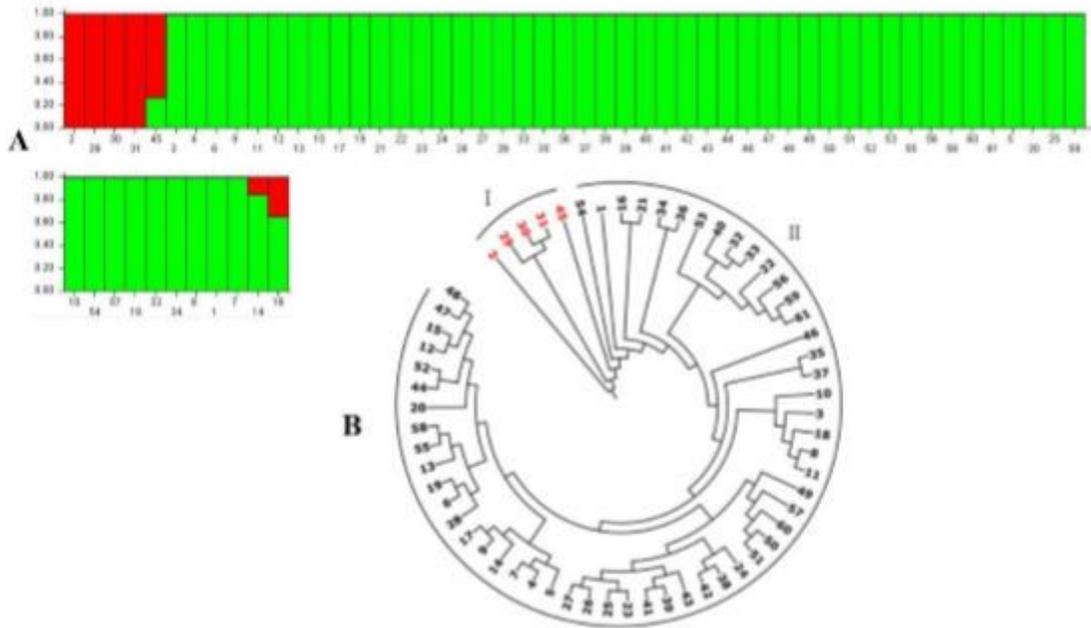


Figure 5

The UPGMA phylogenetic tree of the 61 accessions based on SSR markers. The phylogenetic analysis showed 61 accessions were classified into two groups: group I and II (A). Population structure of 61 accessions in *C. pepo* by Model-based analysis. Scale of Y axis represents the percent of genetic components, and the X axis represents the different materials. Group I and II corresponds to the predefined phylogenetic tree (B).