Genome Wide SSR Development and Their Application in Genetic Diversity Analysis in Wax Gourd

Qianmei Hu

Henan Agricultural University

Haiping Wang

Chinese Academy of Agricultural Sciences

Biao Jiang

Guangdong Academy of Agricultural Sciences

Huayu Zhu

Henan Agricultural University

Xiaoming He

Guangdong Academy of Agricultural Sciences

Pengyao Song

Henan Agricultural University

Jiangping Song

Chinese Academy of Agricultural Sciences

Sen Yang

Henan Agricultural University

Junjun Shen

Northwest Agriculture and Forestry University College of Resources and Environment

Zheng Li

Northwestern Agriculture and Forestry University

Hu Jianbin

Henan Agricultural University

Shouru Sun

Henan Agricultural University

Luming Yang (✉ lumingyang@henau.edu.cn)

Henan Agricultural University <https://orcid.org/0000-0001-9614-4655>

Research article

Keywords: Wax Gourd, SSR, Genetic diversity

Posted Date: January 19th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-147921/v1>

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

[Read Full License](#)

Abstract

Background: Wax gourd (*Benincasa hispida* Cong., $2n=2x=24$) is one of the most important winter vegetables of the Cucurbitaceae family. There are only limited markers available for this crop and the draft genome of wax gourd provides a powerful tool for SSR marker development.

Results: In this study, we developed genome-wide SSR markers from wax gourd genome and characterized their distribution and frequency of different motifs and repeats. A total of 52,431 microsatellites from wax gourd genome were identified, of which 39,319 SSR markers were developed. 1,152 non-wax gourd SSR markers were selected from cucumber, melon, watermelon and pumpkin to test their transferability in wax gourd. 580 SSR markers could be transferable in wax gourd, and 42 of them were detected with polymorphic in 11 tested accessions of wax gourd. In addition, 11 good polymorphic transferrable SSR markers and 21 SSR markers of wax gourd were selected to investigate the genetic diversity and population structure of 129 wax gourd accessions. 112 alleles were detected by these 32 SSR markers. The result of population structure showed that the 129 wax gourd accessions were divided into two main populations, and the genetic diversity analysis separated them into two clusters.

Conclusions: The large number of wax gourd SSR markers developed in this study provides a valuable resource for genetic linkage map construction, molecular mapping, and marker-assisted selection (MAS) in wax gourd.

Background

Wax gourd (*Benincasa hispida* Cong., $2n = 2x = 24$) is an important economically horticultural crop of the Cucurbitaceae family which also includes several other important vegetables such as cucumber (*Cucumis sativus*), melon (*Cucumis melo*), watermelon (*Citrullus lanatus*), bottle gourd (*Lagenaria siceraria*) and pumpkin/squash (*Cucurbita spp.*). China and East India were considered to be the origins of wax gourd[1]. Nowadays, wax gourd has been widely distributed in China, India, and many other countries[2], and it has a cultivation history of more than 2000 years in China. The germplasms of wax gourd showed a wide diversity which can be easily seen from morphological variations of fruit shape, size and weight. The fruit weight of wax gourd ranged from less than 1 kg to more than 25 kg, and the shape is varied from oblate to long cylinder. The fruit of wax gourd can be stored for a long time, which make it plays an important role in annual supply and regulating off-seasons of the vegetables[3]. Wax gourd contains abundant Vitamin C and no-fat, and it has high nutritional and medicinal values. Ripe wax gourd juice can be used to cure insanity and epilepsy[4] and possesses anti-ulcer activity[5].

The purpose of research resources is to take full advantage of heterosis in breeding, while the relationship of different germplasms is difficult to determine by simply relying on agronomic traits or geographical proximity. Modern molecular biotechnology is an effective method to study the genetic relationship and diversity of resources[6]. They can facilitate the rapid screening of polymorphic loci for the large number of genotypes. With the development of molecular biotechnology, a variety of molecular

markers were developed. Among those molecular markers, Simple Sequence Repeat (SSR) marker is an ideal marker in various applications, due to many desirable features including easy to use, relative abundance, reproducibility, codominant inheritance and whole genome coverage[7, 8]. For these excellent advantages of SSR markers, it was widely used in many applications such as gene mapping construction, fingerprinting, genetic diversity, population structure analysis and comparative mapping[9–12].

With the rapid development and cost decreasing of next-generation sequencing (NGS) technologies, the whole genomes of several important crops in the Cucurbitaceae family have been sequenced including cucumber[13], melon[14], watermelon[11], bottle gourd[15], pumpkin/squash[16, 17] and wax gourd[18]. However, the genomic SSR markers of wax gourd are still very limited, only a few SSR markers were developed based on the transcriptome sequences[3], which greatly limits their application in many genetic studies. Up to now, only a high-density genetic map[19], transcriptome sequences for several tissues[3], and a small number of genomic fragments[3, 20] have been developed in wax gourd. Abundant SSR markers have been widely used in many other crops for numerous diverse studies. For instance, the genome wide SSR markers have been developed and applied in watermelon[21], melon[12] and cucumber[22]. The genome wide transferrable SSR markers have also been identified in cucurbit species by comparative genome analysis between cucumber, melon and watermelon[12, 21]. Therefore, it will be valuable to develop whole genome SSR markers of wax gourd for its genetic diversity analysis and genetic mapping.

In the present study, we identified and characterized the distribution and frequency of different microsatellite motifs in the wax gourd genome. In addition, a total of 1,152 non-wax gourd SSR markers were selected to test their transferability in wax gourd. Finally, 11 good polymorphic transferrable SSR markers and 21 SSR markers developed from wax gourd genome were selected and applied in a collection of 129 wax gourd accessions to investigate their genetic diversity and population structure.

Results

The transferability of non-wax gourd SSR marker in wax gourd. Eleven wax gourd accessions with high morphological diversity were selected to test the transferability of 1,152 SSR markers developed from other crops of the Cucurbitaceae family. 580 of them had amplification products in the tested wax gourd accessions, and each of them had clear bands in at least five of these accessions. Of them, the SSR markers developed from watermelon had the highest transferability, and 170 of 288 (59.03%) were transferrable in wax gourd, followed by 153 from melon, 142 from pumpkin and 115 from cucumber. We also checked the polymorphism of 580 SSR markers in eleven wax gourd accessions, and 42 of them were polymorphic in different accessions of wax gourd with 2, 15, 5 and 20 from cucumber, melon, watermelon and pumpkin, respectively (Table 1).

Table 1
The numbers of transferrable and polymorphic SSR markers in wax gourd.

Origins	Watermelon	Melon	Cucumber	Pumpkin	Total
SSR markers	288	288	288	288	1152
Transferable markers	170	153	115	142	580
Polymorphic markers	5	15	2	20	42

Genome wide SSR markers development in wax gourd. The analysis of transferrable markers from non-wax gourd genomes showed a low transferability of these markers, which is far from enough in the genetic study of wax gourd. Therefore, we developed the whole genome wide SSR markers from wax gourd draft genome assembly of B227. A total of 52,431 microsatellite loci were identified from wax gourd draft genome. The total sequence length of all microsatellites accounted for 0.13% of the whole genome, with an average of 55 SSR/Mb. Among different repeat types, the dinucleotides were the most common type accounting for 41.19% of the total SSR loci discovered, followed by trinucleotides (16.71%), while octonucleotides were the least frequent repeat type (3.12%) (Table 2). The SSR motif distribution with regarded to repeat numbers has also been investigated. The microsatellite frequency was decreased as the number of repeat units increased, which was more obvious for longer SSR motifs (Fig. 1). For example, the mean number of repeat motifs in heptanucleotides and octonucleotides were 3.16 and 3.14 respective, and the number of microsatellites were 6,170 and 1,637 respective (Table 2).

Table 2
Distribution of different nucleotide repeats in the wax gourd genome

Motif length	Number of loci identified	Frequency (%)	Mean repeat number	Number of loci primers designed	Percentage SSRs suitable for primer design (%)
Di -	21,594	41.19	11.63	16,339	75.66
Tri -	8,759	16.71	9.36	7,060	80.6
Tetra -	6,154	11.74	5.72	4,513	73.33
Penta -	6,083	11.6	4.26	4,809	79.06
Hexa -	2,034	3.88	4.32	1,761	86.58
Hepta -	6,170	11.77	3.16	4,868	78.9
Octo -	1,637	3.12	3.14	1,356	82.83
Tota	52,431	100		40,706	77.64

Furthermore, the repeat motifs for each type of SSRs identified in the wax gourd genome were also examined. We found that some nucleotide motifs were more prevalent than others. For example, the AT motif was dramatically overrepresented in dinucleotide motifs, and it was also the most frequent motif in

the entire wax gourd genome accounting for 35.8% of the total SSR loci discovered. Similarly, the AAT, AAAT, AAAAT, AAAAAG, AAAAAAT, and AAAAAAAG were the most abundant repeats types in each class (Additional Fig. S 1). The frequency and distribution of SSR in each chromosome showed that the number of microsatellite loci was positively correlated with their chromosome size (Additional Table S 4, Fig. 2). The largest number of microsatellites were detected on chromosome 01 (5,715), followed by chromosome 08 (5,391), and the least SSR number was found on chromosome 07 (2,880). The SSR density near the centromeres is generally low, and there is also a low SSR density at the end of some chromosomes. The sequences containing microsatellite loci were screened for PCR primer design using Primer 3, and 50,298 SSR loci contained suitable flanking sites for SSR primer design. Finally, we designed 39,319 SSR primers with some SSR loci included in the same primers as compound SSRs. The exact positions of these SSRs in the wax gourd chromosomes, as well as information on repeat motifs, expected PCR product size are presented in Additional Table S 5.

The colors from blue to red indicate a gradual increase in the density of SSR markers

Genetic diversity of different phenotypes in wax gourd. We collected 8 phenotypes of 129 wax gourd accessions, and found that the range of coefficient of variation of them was from 0.11 to 0.40 (Table 3). The 129 wax gourd accessions showed relatively narrow variation in eight quantitative traits. The trait with the maximum coefficient of variation was fruit weight (0.40), followed by fruit length (0.28) and fruit diameter (0.20), while the trait with the minimum coefficient of variation was leaf length (0.11). The results showed that the coefficient of variation for fruit related traits was higher than that of all the leaf related traits, indicating that fruit related traits have a higher potential for improvement from these germplasms.

In this study, we also performed a correlation analysis for some related traits. The results showed that there was a high positive correlation between fruit related traits, and the same tendency was also observed among leaf related trait, while the correlation relationship between leaf traits and fruit traits was very weak. Among them, the positive correlation between leaf length and leaf width was the highest (0.92), followed by single fruit weight and flesh thickness (0.69) (Fig. 3).

Table 3
Analysis of variation of quantitative characters for 129 wax gourd germplasms

Trait	Mean	Max.	Min.	Range	S	CV(%)
Fruit length (cm)	41.38	67.00	2.70	64.30	11.55	0.28
Fruit diameter (cm)	19.97	36.50	2.27	34.23	3.90	0.20
Flesh thickness (cm)	4.32	6.50	2.20	4.30	0.82	0.19
Fresh weight of single fruit (kg)	9.37	20.40	2.59	17.81	3.79	0.40
Leaf length (cm)	13.81	17.50	8.34	9.16	1.47	0.11
Leaf width (cm)	19.32	26.05	12.50	13.55	2.51	0.13
Petiole length (cm)	16.61	25.72	10.14	15.58	2.85	0.17
Petiole diameter (cm)	7.32	9.85	3.33	6.51	0.95	0.13

FD (Fruit diameter), (Flesh thickness), FW (Fruit weight of single fruit), FL (Fruit length), LD (Leaf diameter), LL (Leaf length), PL (Petiol length), PD (Petiole diameter).

Genetic diversity analysis of wax gourd accessions. To investigate the genetic diversity of different wax gourd germplasm, 19 SSR markers developed from transcriptome of wax gourd and 48 SSR markers developed from draft genome of wax gourd in this study were used to evaluate their polymorphism in 11 tested wax gourd accessions. 21 of them were identified with good polymorphism, and they were selected together with another 11 non-wax gourd SSR markers in the genetic diversity analysis of 129 wax gourd accessions. The information of those 32 SSR markers was list in Additional Table S6. Totally, 112 alleles were detected by these 32 SSR markers (Table 4). The number of different alleles (N_a) for each marker ranged from 2 to 9, with an average of 3.5. The effective number of alleles ranged from 1.142 to 6.684 with an average of 2.060 per locus. Shannon's information index (I) for each marker ranged from 0.265 to 1.996, with an average of 0.750. Observed heterozygosity (H_o) for each marker ranged from 0.009 to 0.561, with an average of 0.230. Expected heterozygosity (H_e) for each marker ranged from 0.125 to 0.854, with an average of 0.427. Polymorphism Information Content (PIC) value ranged from 0.118 to 0.832, with an average of 0.370 per locus.

Table 4
Genetic diversity characteristics of 129 wax gourd accessions using 32
SSR markers

Locus	<i>Na</i>	<i>Ne</i>	<i>I</i>	<i>Ho</i>	<i>He</i>	<i>PIC</i>
1C096	2	1.405	0.463	0.240	0.289	0.247
1C189	7	4.146	1.577	0.457	0.762	0.721
2C053	5	1.486	0.602	0.233	0.328	0.289
4C020	6	4.068	1.482	0.320	0.757	0.711
5C190	3	2.629	1.022	0.320	0.622	0.542
8C070	3	2.079	0.776	0.516	0.521	0.403
9C051	2	1.231	0.335	0.116	0.188	0.170
9C071	5	1.633	0.802	0.172	0.389	0.365
9C255	2	1.434	0.480	0.217	0.304	0.257
10C006	3	1.142	0.267	0.116	0.125	0.118
11C204	2	1.803	0.637	0.323	0.447	0.346
11C239	5	2.341	1.126	0.192	0.575	0.532
12C147	7	4.208	1.582	0.258	0.765	0.725
BhSSR33868	2	1.160	0.265	0.099	0.138	0.128
BhSSR22321	6	1.742	0.857	0.017	0.428	0.393
BhSSR24640	3	1.805	0.674	0.018	0.448	0.354
BhSSR06554	2	1.930	0.675	0.181	0.484	0.366
BhSSR18918	2	1.677	0.593	0.561	0.405	0.322
BhSSR06950	9	6.684	1.996	0.266	0.854	0.832
BhSSR07727	4	1.743	0.766	0.009	0.428	0.381
BhSSR21714	3	1.925	0.718	0.286	0.483	0.376
SSR17481	3	2.057	0.781	0.419	0.516	0.404
RGA-UW084790	3	1.716	0.749	0.295	0.419	0.379
CmSSR24466	2	1.448	0.488	0.242	0.311	0.262
CmSSR19665	3	1.453	0.588	0.124	0.313	0.287
CmSSR17132	4	1.173	0.347	0.094	0.148	0.143

CmSSR24836	2	1.830	0.646	0.211	0.455	0.351
CmSSR24993	2	1.231	0.335	0.178	0.188	0.170
CmaSSR007847	2	1.973	0.686	0.055	0.495	0.372
CmaSSR005934	4	1.652	0.645	0.508	0.396	0.329
CmaSSR014949	2	1.319	0.406	0.063	0.243	0.212
WMSSR20840	2	1.797	0.636	0.244	0.445	0.345
Mean	3.5	2.060	0.750	0.230	0.427	0.370

Na number of different alleles, *Ne* number of effective alleles, *I* Shannon's information index, *Ho* Observed Heterozygosity, *He* Expected Heterozygosity, *PI*C Polymorphism Information Content

Population structure and genetic diversity analysis of wax gourd germplasm. The population structure of 129 wax gourd accessions was analyzed using a model-based software STRUCTURE that employs Bayesian assignment. Evanno's correction method[23] was applied, which showed a clear peak at $K=2$ (Fig. 4A), and the 129 wax gourd accessions were divided into two main groups (Pop I and Pop II) based on their multi-locus genotypes (Fig. 4B). The group Pop I contained 47 accessions, and most of them (31) were collected from the southern coastal provinces of China. The group Pop II contained 82 accessions, and most of them (71) belonged to Yellow River Basin and Yangtze valley (Additional Table S 1).

The fingerprinting data of 129 wax gourd accessions were used to construct a dendrogram using neighbor-joining method in MEGA 6, and these accessions were classified into 2 major clusters, named cluster I and cluster II, respectively (Fig. 5). The cluster I could be further divided into two sub-clusters, I A and I B. The subcluster I A contained 44 accessions, including 14 accessions from Southern China and 30 from Northern China. The subcluster I B contained 19 accessions, with 4 from Northern China and 15 from Southern China. The cluster II had 24 accessions with 6 from Northern China and 18 from Southern China. The remaining materials are distributed in 8 small clusters. It is obvious that the materials from Northern China were grouped together and those from Southern China were grouped together (Fig. 5).

Discussion

With the rapid development of sequencing technologies, the genome sequence of many field crops and horticultural crops have been completed[11, 13, 24–26]. These available genomic sequences are valuable resources for SSR development, and genome wide identification of SSR have been investigated in many plant species[12, 22]. SSR marker has been widely used in many plants species for genetic diversity analysis such as rice[27], wheat[28], maize[29] and cucumber[10]. Though the genome sequence of wax gourd had been sequenced recently, the genome wide SSR markers have not been developed. The lack of sufficient molecular markers has become a major challenge and restricted the development of many studies, such as genetic diversity analysis, fine mapping and genome wide association analysis. SSR markers are conserved among closely related species and genus, thus SSR markers have been proved to

be useful in cross-species transformation[30–32]. Cross-species transferability of SSR markers offered an easy, time and cost effective way to develop SSR makers in such species from related species whose SSR markers have been developed[33]. They are not only useful in the genetic diversity analysis and map construction in closely related species, but also provide important genetic information for comparative genomics[11, 21, 34]. Therefore, the transferability of SSR markers has been widely used in the family of Poaceae[32], Leguminosae[31], Vitaceae[35], and Cucurbitaceae[30, 36, 37]. However, the transferability of SSR markers in wax gourd from other cucurbit species is still large unknown. In the present study, the transferability of SSR markers from cucumber, melon, watermelon, and pumpkin were tested in wax gourd by selecting 288 SSR markers developed from each of these species. The number of transferability of SSR markers in wax gourd was 170, 153, 142 and 115 from watermelon, melon, pumpkin and cucumber, respectively. The transferability of SSR markers are usually higher between the evolutionary more closely related species[21]. Among these five cucurbit species, the wax gourd is more closely related to watermelon[38], and the highest number of transferable SSR markers were also developed from watermelon which were consist with their evolutionary genetic relationship.

Although the number of transferrable markers may be higher, the proportion of polymorphism may be very lower. In our study, the transferrable SSR markers of non-wax gourd species showed a low polymorphism in wax gourd germplasms. It is far from enough to meet the needs of map construction and others genetic research. With the draft genome sequence of wax gourd available, we identified 52,431 microsatellites from wax gourd B227 genome assembly with a frequency of 55 SSR/Mb in this study (Table 2). The number of microsatellites and their density identified in our study was lower than that in cucumber (552 SSR/Mb) and Arabidopsis (371 SSR/Mb)[22], watermelon (111 SSR/Mb)[21], and melon (109 SSR/Mb)[12]. One main reason for these differences was due to the search parameters used for detection of microsatellites. For example, different repeat types (mononucleotides to pentanucleotides vs. mononucleotides to octanucleotides) of different minimum lengths (12 vs. 18 bp) were searched using the same or different software. In the present study, we analyzed the distribution and frequency of microsatellites with motifs of 2–8 bp long and minimum lengths of 18 bp or minimum of three repeat units in wax gourd genome (Fig. 1). The criterion we used was according to the fact that polymorphism levels and mutation rate correlate positively with the number of repeat units. Different studies have revealed that high numbers of repeats especially for dinucleotides and trinucleotides in long microsatellites are more likely to be polymorphic as compared to shorter one because of higher rate of DNA replication slippage[39, 40].

The differential SSR distribution has been reported between intronic and intergenic regions, and different chromosomes, and different species have different frequencies of SSR types and repeat units[41–43]. Frequency analysis of various nucleotide repeats in wax gourd revealed that dinucleotide repeats were the most abundant SSRs followed by trinucleotide, tetranucleotide, pentanucleotide, heptanucleotide, hexanucleotide, and octonucleotide repeats (Fig. 1 and Table 2). This was different from the trend in other species. For example, the tetranucleotide repeats were the most abundant in cucumber, *Medicago truncatula*, *Populus trichocarpa*, and *Vitis vinifera*, and the trinucleotide repeats were the most abundant in *Glycine max*, *Arabidopsis thaliana*, *Oryza sativa*, and *Sorghum bicolor*[22]. Overall, the AT rich motifs

such as AT and AAT were the predominant SSR repeat types in each class in wax gourd, representing 41.2% and 16.7% in dinucleotide repeats and trinucleotide repeats, respectively. Conversely, GC-rich repeat SSR motifs were very rare in all the nucleotide repeats. This result is consistent with other studies indicating that genomic SSRs with GC-rich repeats are rare in plant species[44, 45]. The frequency and distribution of different SSR type in different chromosomes revealed that the frequency of microsatellite loci was positively correlated with the chromosome size in wax gourd. This was different from the trend in melon and water melon[12, 21].

SSR markers have been used in the genetic analysis of many horticultural crops[12, 21, 46–48]. Though there are a few SSR markers of wax gourd developed from the transcriptome[3], the genetic diversity of wax gourds has been rarely analyzed in previous studies. In this study, 32 SSR markers were used for inferring population structure and genetic diversity analysis of 129 wax gourd accessions. 112 alleles were detected in the 129 wax gourd accessions using the 32 markers and the number of different alleles (N_a) for each marker ranged from 2 to 9, and their PIC values ranged from 0.118 to 0.832 with an average of 0.370. The average of alleles and PIC values were lower than studies in bottle gourd[48], radish[46] and watermelon[47]. The lower levels of polymorphism may be caused by the narrow genetic background of 129 wax gourd accessions, which were all collected from China. This was further confirmed by population structure and genetic diversity analysis. The 129 wax gourds were divided into two populations and clusters (Figs. 4 and 5), and some accessions from different regions were mixed, suggesting that they have similar genetic background.

Conclusions

In our study, we identified 52,431 microsatellites from wax gourd genome and developed 39,319 SSR markers from them loci. The distribution and frequency of different motifs and repeats was also characterized on different chromosomes. The large number of wax gourd SSR markers developed in this study provides a valuable resource for genetic linkage map construction, molecular mapping, and marker-assisted selection (MAS) in wax gourd. Furthermore, the transferability of SSR markers developed from cucumber, melon and watermelon was also validated and compared in wax gourd. The genetic diversity and population structure of 129 wax gourd germplasm collected from different provinces of China was investigated, which showed the distribution and clustering of Chinese wax gourd are closely related to the distribution of water system in China. These data have improved our understanding of domestic wax gourd germplasm resources and conducive to the effective research, utilization and introduction of germplasm resources.

Methods

Plant materials. A total of 129 wax gourd accessions were used in this study for genetic diversity analysis, which were collected from 23 different provinces of China. These accessions were provided by the national germplasm repository for vegetatively propagated vegetables. Of them, 24 accessions were selected from Henan province, followed by 16 from Hunan, 14 from Shandong, 12 from Jiangsu, 10 from

Fujian, and the all the remaining provinces have less than 10 accessions. The origin information of 129 wax gourd accessions was list in Additional Table S 1. All the wax gourds were grown in the open field at Maozhuang Research Station of Henan Agricultural University (Zhengzhou, China). The phenotypes of eight traits were collected from 129 wax gourd accessions including fruit diameter, fruit length, flesh thickness, single fruit weight, leaf diameter, leaf length, petiole length and petiole diameter. These phenotypes were collected from five plants of each accession. The statistical formula of coefficient of variation: $CV = SD / X$, and the correlation were calculated using R 3.5.1.

Non-wax gourd SSR markers selection. In order to test the transferability of SSR markers from other species of the Cucurbitaceae family in wax gourd, 1,152 SSR markers were selected from cucumber, melon, watermelon and pumpkin with 288 SSR markers from each crop. These SSR markers of four species were selected based on their physical position in the genome assembly. In cucumber, an average of 41 markers were selected from each of the seven cucumber chromosomes. In pumpkin, 288 SSR markers were selected from all the 20 chromosomes with at least three markers from each of the 20 chromosomes. In watermelon, 288 SSR markers from 2–11 chromosomes, it is evenly distributed across each chromosome. In melon, 288 SSR markers were selected from seven chromosomes, they are mainly come from chromosome 7–12, except for three markers from the end of chromosomes 6. The detail information for these SSR markers were provided in Additional Table S 2. All these SSR markers were grouped into non-wax gourd SSR markers in this study. Furthermore, 19 SSR markers of wax gourd (Additional Table S 3) provided by Guangdong Academy of Agricultural Sciences were also used as control in genetic diversity analysis.

SSR identification and primer design in wax gourd genome. The wax gourd draft genome was downloaded from Cucurbit Genomics Database (<http://cucurbitgenomics.org/>). In order to develop a higher polymorphism SSR platform for future study, the parameter of microsatellite identification in this study was from 2- to 8-bp motifs, and mononucleotides were not considered due to the difficulty of distinguishing bona fide microsatellites from sequencing or assembly error. DNA sequences were searched for both perfect and compound microsatellites, with a basic motif of 2–8 bp, using the computer program MISA (Microsatellite identification tool) [49]. Repeats with a minimum length of 18 (for di- to tetranucleotides), 20 (for pentanucleotides), 24 (for hexanucleotides), 21 (for heptanucleotides), and 24 bp (for octanucleotides) were recorded. The physical positions of the SSRs found in the chromosomes were also recorded, and oligonucleotide primers were designed for the genomic sequence flanking these SSRs using Primer 3 (v. 1.1.4) software[50]. Primers were designed to generate amplicons of 100–300 bp in length with the following minimum, optimum and maximum values for Primer 3 parameters: primer length (bp) 18-20-24 and Tm (°C) 50-55-60. Other parameters used the default program values.

DNA extraction and PCR amplification. The unexpanded young leaves from each accession were collected into 2.0 mL microcentrifuge tubes, lyophilized in a freeze dryer, and ground into fine powder. Genomic DNA was extracted with a modified cetyl trimethylammonium bromide (CTAB) method[51], and the quality of DNA was quantified by spectrophotometer and further checked on agarose gel. Each polymerase chain reaction (PCR) contained 25 ng template DNA, 0.5 μ M each of forward and reverse

primers, 0.2 mM dNTPs mix, 0.5 unit of Taq DNA polymerase and 1 × PCR buffer in a total volume of 10.0 µl. The amplification was carried out at initial denaturing step at 94 °C for 4 min followed by 30 cycles of 94 °C for 20 sec, 58 °C for 45 sec and 72 °C for 1 min. In the last cycle, primer extension was performed at 72 °C for 10 min and storage at 4 °C till electrophoresis. The PCR products were size-fractionated in a 9% polyacrylamide gel. The 100-bp DNA ladder was used as molecular size marker. After gel electrophoresis, band patterns were visualized with silver staining, and gel images were taken with a digital camera.

Data analysis. To test the transferrable SSR markers of non-wax gourd crops, we used the DNA template of crops as positive control in which the SSR markers were developed and collected. In addition, 11 wax gourd germplasms with high diversity in morphological difference were selected to test the polymorphism of SSR markers. Furthermore, 12 wax gourd accessions with high diversity in morphological difference were selected to test the polymorphism of 67 SSR markers developed from wax gourd genome.

The polymorphic SSR markers were manually scored as binary data with presence as “1” and absence as “0”. The observed (N_a) and effective (N_e) number of alleles, Shannon’s information index (I), and levels of observed (H_o) and expected (H_e) heterozygosity were calculated by Popgen 32. Polymorphic information content (PIC) for molecular markers was calculated as $PIC = 1 - \sum P_{ij}^2$, where P_{ij} is the frequency of the i th allele for the j th SSR locus[52]. The genotypic distance matrix were conducted using the GeneA1Ex 6.5[53] and the neighbor-joining method in the software MEGA 6[54] was used to construct the dendrogram on the basis of the distance matrix.

The population structure of 138 wax gourd accessions was inferred using STRUCTURE 2.3.4[55]. Several population numbers (from $K = 1$ to 10) were tested by the software to identify the highest ΔK who represented the true value of K [23]. The option of correlated allele frequencies were selected, a burn-in period of 50,000 steps, and 100,000 Markov Chain Monte Carlo (MCMC) replicates; each run was replicated 10 times to ensure consistency of results.

Abbreviations

NGS: Next-generation sequencing; SSR: Simple sequence repeats; N_a : Number of different alleles; N_e : Number of effective alleles; I : Shannon’s information index; H_o : Observed Heterozygosity; H_e : Expected Heterozygosity; PIC: Polymorphism Information Content; MISA: Microsatellite identification tool; PCR: Polymerase Chain Reaction.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of date and materials

All of the data generated or analyzed during this study are included in this published article (and its supplementary information files).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

HQM, ZHY, SSR and YLM conceived and designed the experiments and YLM responsible for funding acquisition. All authors were involved in defining the experimental strategy. HQM, JB, WHP and SPY performed the experiments. HQM and JS processed and analyzed data. All authors contributed to writing and revising the manuscript. All the authors read and approved the final manuscript.

Acknowledgements

This work was supported by the Project for Scientific and Technological Activities of Overseas Students of Henan Province and the Zhongyuan Youth Talent Support Program (ZYQR201912161).

Author details

¹ College of Horticulture, Henan Agricultural University, 63 Nongye Road, Zhengzhou, 450002, China.

² Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, 12 Nandajie, Zhongguancun, Haidian District, Beijing 100081, China.

³ Vegetable Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, Guangdong 510640, China.

⁴ College of Horticulture, Northwest A&F University, Yangling, China.

References

1. Rubatzky VE, Yamaguchi M. World vegetables: principles, production, and nutritive values: Chapman & Hall, International Thomson Pub.; 1997.
2. Chopra RN. Glossary of Indian medicinal plants. *Quarterly Review of Biology* 1956.
3. Jiang B, Xie D, Liu W, Peng Q, He X. De Novo Assembly and Characterization of the Transcriptome, and Development of SSR Markers in Wax Gourd (*Benincasa hispida*). *Plos One* 2013; 8.
4. Ramesh M, Gayathri V, Rao AVNA, Prabhakar MC, Rao CS. Pharmacological actions of fruit juice of *Benincasa hispida*. *Fitoterapia*. 1989;60(3):241–7.

5. Grover JK, Adiga G, Vats V, Rathi SS. Extracts of *Benincasa hispida* prevent development of experimental ulcers. *J Ethnopharmacol*. 2001;78(2–3):159–64.
6. Thormann CE, Ferreira ME, Camargo LEA, Tivang JG, Osborn TC. Comparison of RFLP and RAPD markers to estimating genetic relationships within and among cruciferous species. *Theoretical Applied Genetics*. 1994;88(8):973–80.
7. Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK. Microsatellite markers: an overview of the recent progress in plants. *Euphytica*. 2011;177(3):309–34.
8. Varshney RK, Graner A, Sorrells ME. Genic microsatellite markers in plants: features and applications. *Trends Biotechnol*. 2005;23(1):48–55.
9. Bruford MW, Wayne RK. Microsatellites and their application to population genetic studies. *Curr Opin Genet Dev*. 1993;3(6):939–43.
10. Lv J, Qi J, Shi Q, Shen D, Zhang S, Shao G, Li H, Sun Z, Weng Y, Shang Y. Genetic Diversity and Population Structure of Cucumber (*Cucumis sativus* L.). *Plos One*. 2012;7(10):e46919.
11. Guo S, Zhang J, Sun H, Salse J, Lucas WJ, Zhang H, Zheng Y, Mao L, Ren Y, Wang Z. The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nature Genetics* 2013.
12. 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.
13. Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P, et al. The genome of the cucumber, *Cucumis sativus* L. *Nature Genetics* 2009; 41(Suppl 2).
14. Garcia-Mas J. The genome of melon (*Cucumis melo* L.). *Proceedings of the National Academy of Sciences of the United States of America* 2012.
15. Wu S, Shamimuzzaman M, Sun H, Salse J, Sui X, Wilder A, Wu Z, Levi A, Xu Y, Ling KS. The bottle gourd genome provides insights into Cucurbitaceae evolution and facilitates mapping of a Papaya ringspot virus resistance locus. *Plant Journal* 2017.
16. Sun H, Wu S, Zhang, Guoyu J. Chen, Guo, Shaogui. Karyotype Stability and Unbiased Fractionation in the Paleo-Allotetraploid *Cucurbita* Genomes. *Molecular Plant* 2017.
17. Montero-Pau J, Blanca J, Bombarely A, Ziaresolo P, Caizares J. De-novo assembly of zucchini genome reveals a whole genome duplication associated with the origin of the *Cucurbita* genus. *Plant Biotechnology Journal* 2018; 16(6).
18. Xie D, Xu Y, Wang J, Liu W, Zhang Z. The wax gourd genomes offer insights into the genetic diversity and ancestral cucurbit karyotype. *Nat Commun*. 2019;10(1):5158.
19. Jiang B, Liu W, Xie D, Peng Q, He X, Lin YE, Liang Z. Additional file 2: of High-density genetic map construction and gene mapping of pericarp color in wax gourd using specific-locus amplified fragment (SLAF) sequencing. 2015.

20. Bhavé MR, Gupta VS, Ranjekar PK. Arrangement and size distribution of repeat and single copy DNA sequences in four species of Cucurbitaceae. *Plant Systematics Evolution*. 1986;152(3–4):133–51.
21. 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(1):557.
22. 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(1):569.
23. Evanno GS, Regnaut SJ, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol*. 2010;14(8):2611–20.
24. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, et al. The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science*. 2015;326(5956):1112–5.
25. Consortium TG. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 2012.
26. Jia J, Zhao S, Kong X, Li Y, Zhao G, Ai E. *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature*. 2013;496(7443):91–5.
27. Ashraf H, Husaini AM, Ashraf Bhat M, Parra G, Khan S, Ganai NA. SSR based genetic diversity of pigmented and aromatic rice (*Oryza sativa* L.) genotypes of the western Himalayan region of India. *Physiology Molecular Biology of Plants*. 2016;22(4):1–9.
28. Mostafa R, Ghader, Mirzaghaderi, Salar, Shaaf, Hedyeh, Badakhshan. SSR assessment of the genetic diversity of emmer wheat with emphasis on Iranian landraces (*Triticum dicoccon* Schrank). *Genetic Resources & Crop Evolution* 2016.
29. Mehta B, Hossain F, Muthusamy V, Baveja A, Zunjare R, Jha SK, Gupta HS. Microsatellite-based genetic diversity analyses of sugary1-, shrunken2- and double mutant- sweet corn inbreds for their utilization in breeding programme. *Physiology Molecular Biology of Plants*. 2017;23(2):411–20.
30. Ghebretinsae AG, Thulin M, Barber JC. Relationships of cucumbers and melons unraveled: molecular phylogenetics of *Cucumis* and related genera (Benincaseae, Cucurbitaceae). *Am J Bot*. 2007;94(7):1256–66.
31. Wang LX, Cheng XZ, Wang SH, Liu CY, Liang H. Transferability of SSR Markers from Adzuki Bean into Mungbean. *Acta Agronomica Sinica*. 2009;35(5):816–20.
32. Daniele T, Marco, Maccaferri, Peter, de, Heer, Anker, Sørensen, Silvia. High-throughput SNP discovery and genotyping in durum wheat (*Triticum durum* Desf.). *Theoretical & Applied Genetics* 2011.
33. Barbará T, Palma-Silva C, Paggi GM, Bered F, Fay MF, Lexer C. Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Mol Ecol*. 2010;16(18):3759–67.
34. Yang L, Koo DH, Li Y, Zhang X, Luan F, Havey MJ, Jiang J, Weng Y. Chromosome rearrangements during domestication of cucumber as revealed by high-density genetic mapping and draft genome assembly. *Plant J*. 2012;71(6):895–906.

35. Rossetto M, McNally J, Henry RJ. Evaluating the potential of SSR flanking regions for examining taxonomic relationships in the Vitaceae. *Theoretical Applied Genetics*. 2002;104(1):61–6.
36. Verma VK, Behera TK, Munshi AD, Parida SK, Mohapatra T. Genetic diversity of ash gourd [*Benincasa hispida* (Thunb.) Cogn.] inbred lines based on RAPD and ISSR markers and their hybrid performance. *Sci Hortic*. 2007;113(3):231–7.
37. Watcharawongpaiboon N, Chunwongse J. Development and Characterization of Microsatellite Markers from an Enriched Genomic Library of Cucumber (*Cucumis sativus*). *Plant Breeding* 2008; 127(1).
38. Sebastian P, Schaefer H, Telford IRH, Renner SS. Cucumber (*Cucumis sativus*) and melon (*C. melo*) have numerous wild relatives in Asia and Australia, and the sister species of melon is from Australia. *Proceedings of the National Academy of Sciences of the United States of America* 2010; 107(32).
39. Weber LJ. Informativeness of human (dC-dA)_n.(dG-dT)_n polymorphisms. *Genomics*. 1990;7(4):524–30.
40. Yves V, Jaqueth JS, Yoshihiro M, Smith OS, Beavis WD, Smith JSC, John D. Rate and pattern of mutation at microsatellite loci in maize. *Molecular Biology & Evolution* 2002(8):1251–1260.
41. Bian Y, Ballington J, Raja A, Brouwer C, Reid R, Burke M, Wang X, Rowland LJ, Bassil N, Brown A. Patterns of simple sequence repeats in cultivated blueberries (*Vaccinium* section *Cyanococcus* spp.) and their use in revealing genetic diversity and population structure. *Mol Breeding*. 2014;34(2):675–89.
42. Shi J, Huang S, Zhan J, Yu J, Wang X, Wei H, Liu S, Liu G, Wang H. Genome-Wide Microsatellite Characterization and Marker Development in the Sequenced Brassica Crop Species. *Dna Research An International Journal for Rapid Publication of Reports on Genes & Genomes* 2013(1):53–68.
43. Tan M, Wu K, Wang L, Yan M, Zhao Z, Xu J, Zeng Y, Zhang X, Fu C, Xue J. Developing and characterising *Ricinus communis* SSR markers by data mining of whole-genome sequences. *Mol Breeding*. 2014;34(3):893–904.
44. Wang Z, Weber JL, Zhong G, Tanksley SD. Survey of short tandem DNA repeats. *Theoretical Applied Genetics*. 1994;88(1):1–6.
45. Tangphatsornruang S, Somta P, Uthaisaipaisanwong P, Chanprasert J, Sangsrakru D, Seehalak W, Sommanas W, Tragoonrung S, Srinives P. Characterization of microsatellites and gene contents from genome shotgun sequences of mungbean (*Vigna radiata* (L.) Wilczek). *Bmc Plant Biology*. 2009;9(1):137.
46. Lee ON, Han YP. Assessment of genetic diversity in cultivated radishes (*Raphanus sativus*) by agronomic traits and SSR markers. *Scientia Horticulturae* 2017.
47. Mashilo J, Shimelis H, Odindo AO, Amelework B. Genetic diversity and differentiation in citron watermelon [*Citrullus lanatus* var. *citroides*] landraces assessed by simple sequence repeat markers. *Sci Hortic*. 2017;214:99–106.
48. Mashilo J, Shimelis HA, Odindo AO, Amelework BA. Genetic differentiation of bottle gourd [*Lagenaria siceraria* (Molina) Standl.] landraces assessed by fruit qualitative traits and simple sequence repeat

- markers. *Sci Hortic.* 2017;216:1–11.
49. Thiel T, Michalek W, R., VarshneyA., Graner. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Tag Theoretical & Applied Genetics* 2003.
50. S R, H S. Primer3 on the WWW for general users and for biologist programmers. *Methods in molecular biology (Clifton, NJ)* 2000; 132.
51. Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 1980;8(19):4321–5.
52. Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME. Optimizing parental selection for genetic linkage maps. *Genome.* 1993;36(1):181–6.
53. 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.
54. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology Evolution.* 2013;30(12):2725–9.
55. Falush D, Stephens M, Pritchard, Jonathan K. Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics* 2003.

Figures

Figure 1

Distribution of SSR motif repeat numbers and relative frequency in the wax gourd genome. The vertical axis shows the abundance of microsatellites that have different motif repeat number (from 3 to >15), which are discriminated by legends of different colors

Figure 2

The density distribution of SSR on 12 chromosomes of wax gourd. The colors from blue to red indicate a gradual increase in the density of SSR markers

Figure 3

Correlation analysis of eight morphological traits. FD (Fruit diameter), (Flesh thickness), FW (Fruit weight of single fruit), FL (Fruit length), LD (Leaf diameter), LL (Leaf length), PL (Petiol length), PD (Petiole diameter).

Figure 4

Delta K distribution across various clusters (K) as estimated by Structure Harvester (A) and Population structure of 129 wax gourd accessions (B). In B, scale of Y axis represents the percent of genetic components, and the X axis represents the wax gourd accessions. Each grid represents a wax gourd accession

Figure 5

The phylogenetic tree of 150 wax gourd accessions constructed by neighbor-joining (NJ) method. The roman numerals represent different clusters. The color branches represent the accessions collected from different regions.

Supplementary Files

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

- [SupplementaryPicture.docx](#)
- [SupplementaryTables.xlsx](#)