Development and Characterization of Simple Sequence Repeat (SSR) Markers from the Genomic sequence of Sweet potato [*Ipomoea batatas* L. (Lam) cv. Taizhong6]

Hanna Amoanimaa-Dede
Guangdong Ocean University

Jiacheng Zhang
Guangdong Ocean University

Chuntao Su
Guangdong Ocean University

Hongbo Zhu
Guangdong Ocean University

Research article

Keywords: genome sequences, SSR markers, sweet potato, genetic diversity, polymorphism

Abstract

The genetic complexity and the lack of genomic resources in sweet potato make genetic improvement extremely challenging. Simple sequence repeats (SSRs) offer an effective molecular marker technology for molecular-based breeding and for locating important loci in crop plants, but only a few have previously been developed in sweet potato. To explore new SSR markers and accelerate its use in sweet potato, we developed a set of 2,431 primer pairs from 133,727 SSRs identified in the sweet potato genome using the Perl script MISA software. The average frequency was one SSR per 6.26 kb with dinucleotides (38.5%) being the most dominant repeat motif. The main motif types in all repeats were AT/AT, AAT/ATT, A/T, AAAT/ATTT, AAAAT/ATTTT and AAAAAT/ATTTTT accounting for 78.29% of the total SSRs. Out of the 100 randomly selected primer pairs, 50% produced clear bands and amplified 251 alleles. On average, the number of alleles was 5.02 per locus for values ranging from 1 to 13 alleles. The UPGMA cluster analysis grouped the 24 sweet potato materials into four clusters at a similarity coefficient of 0.68 showing no relationship between the genotypes and the geographic sources of germplasm. The SSR markers currently developed will provide valuable genetic resources for germplasm identification, and accelerate studies on genetic diversity in sweet potato and related species.

1.0 Introduction

Sweet potato [*Ipomoea batatas* L. (Lam)] is an essential food crop belonging to the *Convolvulaceae* family. Currently, it is cultivated extensively in over 100 countries worldwide but originally native to Central America [1]. It is of great economic importance due to the excellent supply of dietary fiber, vitamins, minerals, and phenolic compounds [2]. Sweet potato is appreciated greatly for its health-promoting functions such as anti-carcinogenic and cardiovascular disease-preventing properties in addition to its radical-scavenging activity [3]. Sweet potato is a highly heterogeneous auto-hexaploid (2n = 6x = 90) plant with complex genetics and a lack of genomic resources [4] making its genetic and functional analyses extremely difficult. Thus, molecular marker technology through marker-assisted selection can be an effective tool to improve genomic selection in sweet potato and to accelerate breeding progress.

Molecular marker technology plays an essential role in genetic diversity and relationship assessment in plants [5]. The molecular markers generally applied in sweet potato genetic analyses include random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP) and microsatellite or simple sequence repeat (SSR). These markers accelerate cultivar development but its development and application in sweet potato genetic improvement is minimal [6]. Simple Sequence Repeats (SSR) is a short repeat sequence composed of 1–6 bases, widely distributed in the genomes of both prokaryotic and eukaryotic organisms [7]. SSR markers are the most ideal molecular marker for crop improvement due to their allele specificity, high polymorphism, and co-dominant nature [8, 9]. These markers are extensively used in many genetic-based studies.

To a greater extent, SSR markers in sweet potato are derived from expressed sequences mined from public databases [10], genomic libraries [11] and pyrosequencing [12, 13]. However, the development of microsatellite markers based on published genome sequences remains limited in sweet potato. Recently, SSR markers have been used to study the dispersal and origin of sweet potato [14], genetic diversity analysis [15] and construction of genetic linkage maps [16]. Out of the 79 primer pairs designed from small-insert and enriched libraries, 34% were polymorphic among the sweet potato accessions examined [11]. Wang et al. [10] found 7,958 SSR loci in 7,163 ESTs from which 1,060 SSR primer pairs were developed. Among them, 816 primer pairs amplified clear reproducible bands.

In recent years, thousands of microsatellite markers have evolved from investigating transcriptome libraries and expressed sequence tags (ESTs) in sweet potato. Nonetheless, the evolution and application of microsatellite markers from genomic sequences are still limited due to the lack of genomic resources. Though several SSR markers have been

developed and used in sweet potato, their number and availability are low compared to other crops and only a few could amplify or show polymorphisms among the diverse sweet potato varieties [10, 17]. In our opinion, no research has been conducted on the use of genomic sequences to develop SSR markers in sweet potato. In this study, we evaluated the frequency and relative number of SSRs in the sequenced genome of sweet potato (cv. Taizhong6), developed a set of SSR markers from the assembled genomic sequences and assessed the genetic diversity in cultivated sweet potato. These SSR markers will offer new genetic resources for the genetic improvement of sweet potato, adding up to the available resources for analyzing the molecular phylogeny and genetic diversity in sweet potato and related species.

2.0 Materials And Methods

2.1 Plant materials and DNA extraction

A total of 24 sweet potato materials were sampled (Table 1) for this study. The test materials were obtained from the Research Field of Guangdong Ocean University. The genomic sequence of sweet potato cv. Taizhong6 was downloaded from the NCBI database (https://www.ncbi.nlm.nih.gov/). CTAB (cetyltrimethyl ammonium bromide) protocol was employed for the genomic DNA extraction with little modifications [18]. The amplified PCR fragments were verified on 1% agarose gel to test the quality of the DNA.

Table 1: List of 24 sweet potato materials used in the study

No.	Name	Type	Origin
1.	"Guangshu87"	Cultivar	Guangdong, China
2.	"Pushu32"	Cultivar	Guangdong, China
3.	"Ziluolan"	Cultivar	Guangdong, China
4.	"Wulixiang"	Cultivar	Guangdong, China
5.	"Dayehong"	Cultivar	Guangdong, China
6.	"Tianeshu"	Cultivar	Guangdong, China
7.	"Jishu26"	Cultivar	Shandong, China
8.	"Xinong431"	Cultivar	Shaanxi, China
9.	"Qinshu 8"	Cultivar	Shaanxi, China
10.	"Qinshu 5"	Cultivar	Shaanxi, China
11.	"Shangshu19"	Cultivar	Henan, China
12.	"Longshu 9"	Cultivar	Fujian, China
13.	"Xinnianggao"	Cultivar	Fujian, China
14.	"Simon 1"	Cultivar	Brazil
15.	"Taiwanziyang"	Cultivar	Taiwan, China

16.	"Au-1"	Cultivar	Australia
17.	"AU-2"	Cultivar	Australia
18.	"AU-3"	Cultivar	Australia
19.	"Yizi138"	Cultivar	Beijing, China
20.	"Longshu 515"	Cultivar	Fujian, China
21.	"Fushu 18"	Cultivar	Fujian, China
22.	"Guijingzi 8"	Cultivar	Guangxi, China
23.	"Guicai 1"	Cultivar	Guangxi, China
24.	"Zheshu 75"	Cultivar	Zhejiang, China

2.2 SSR mining and primer design

SSRs from the sweet potato genome were identified using the Perl Script MISA (MIcroSAtellite) identification software (http://pgrc.ipk-gatersleben.de/misa)[19]. The minimum consecutive repeat units for mono- to hexanucleotide sequences were 20, 10, 7, 5, 4, and 4 respectively. Primers were designed based on the selected SSR motifs using Primer3 ver. 4.0.0 [20]. The amplified product size was 100–280 bp, primer length ranged between 18–27 bp (optimum; 20 bp) and the minimum interval between two SSR sequences was 100 bp. The CG contents ranged between 20% and 80%, the optimum melting temperature (Tm) was 60 °C (57 °C to 63 °C) and all other parameters were default values.

2.3 SSR primer validation and marker amplification

From the developed primers, 50 primer pairs with a good amplification effect and clear stable bands were used to amplify the genomic DNA of 24 sweet potato materials. The frequency and distribution of all SSRs were analyzed and measured as one SSR/kb of sequence while SSR markers were grouped according to the location of the SSR motifs in the gene. PCR amplification was performed in a reaction volume of 20 µl containing 1.5 µl genomic DNA (20 ng/µl), Taq enzyme 0.4 µl (3 U/µl), 10 × PCR Buffer 2 µl, dNTP 0.2 µl (10 mmol/L), 1 µl each of Forward and Reverse primers (2 µmol/L), sterile double distilled water was added to the final volume [21]. PCR reaction procedure was: pre-

denaturation at 94 °C for 5 min, followed by 33 cycles of denaturation at 94 °C for 30 s, annealing at 55 °C for 30 s, extension at 72 °C for 30 s and a final extension at 72 °C for 5 min using the Takara PCR Thermal Cycler Dice.

2.4 Allele identification and genetic diversity analysis

To identify alleles, the amplified products were identified and separated on a 6% non-denaturing polyacrylamide gel electrophoresis (PAGE). The electrophoresis buffer contained 1 × TBE (100 mM Tris−HCl, 83 mM boric acid, 1 mM Na$_2$EDTA, pH 8.0) [22]. Electrophoresis at 200 V and 100 mA for 120 min was executed after loading of samples. The silver staining method was employed for allele visualization and selection of suitable SSR primers. Markers were scored manually and the polymorphic ones selected for genetic analysis [23].

Each SSR allele for a given primer pair was scored 1 and 0 for samples with and without band respectively, and a "0, 1" binary matrix established. PowerMarker ver.3.25 was used to calculate the number of allele per locus for the genotypes [24].The clustering analysis based on the genetic similarity coefficient was calculated by the unweighted pair group method with arithmetic average (UPGMA) method [25] and a dendrogram generated using the NTSYS software ver. 2.2 [26].

3.0 Results

3.1 Frequency and distribution of SSRs in the sweet potato genome

Based on the sweet potato (cv. Taizhong6) genome sequence, SSRs identified were characterized as mono- to hexanucleotides. The 133,727 SSRs identified had a density of 159.77 SSR/Mb or one SSR per 6.26 kb of sequence on average. Dinucleotides (38.50%), trinucleotides (31.45%) and mononucleotides (12.77%) were the most abundant repeat motif representing 82.71% of the total SSR while the remaining repeat motifs represented 17.29% (Table 2; Additional file 2: Figure A1). The most dominant motifs accounted for 78.29% of all SSRs, among them, AT/AT (81.24%), AC/GT (9.98%) and AG/CT (8.77%) were the most abundant dinucleotide repeats while A/T (98.75%), AAT/ATT (87.17%) and AAAT/ATTT (56.01%) were the most dominant motifs among mononucleotides, trinucleotides and tetra nucleotides respectively (Table 2; Additional file 2: Figure A1). Of all the repeating sequences, AT/AT were the most common representing 31.27% followed by AAT/ATT (27.42%), A/T (12.61%), AAAT/ATTT (4.67%), AC/GT (3.84%), AG/CT (3.38%), and AAAAT/ATTTT (1.7%; Table 2).

Table 2

Distribution of SSRs in the sweet potato genome.

Repeat type	Repeat motif	Number of SSRs	Proportion in all SSRs (%)	Density (SSR/Mb)
Mononucleotide				
	A/T	16859	12.61	20.14
	C/G	213	0.16	0.25
		17072	12.77	20.4
Dinucleotide				
	AT/AT	41822	31.27	49.96
	AC/GT	5136	3.84	6.14
	AG/CT	4516	3.38	5.4
	CG/CG	5	0	0.01
		51479	38.5	61.5
Trinucleotide				
	AAT/ATT	36663	27.42	43.8
	AAC/GTT	2303	1.72	2.75
	AAG/CTT	1668	1.25	1.99
	others	1425	1.07	1.7
		42059	31.45	50.25
Tetranucleotide				
	AAAT/ATTT	6247	4.67	7.46
	ACAT/ATGT	2007	1.5	2.4
	AATT/AATT	844	0.63	1.01
	AAAC/GTTT	789	0.59	0.94
	others	1267	0.95	1.51
		11154	8.34	13.33
Pentanucleotide				
	AAAAT/ATTTT	2374	1.78	2.84
	AAAAC/GTTTT	1369	1.02	1.64
	AATAT/ATATT	680	0.51	0.81
	others	3077	2.3	3.68
		7500	5.61	8.961
Hexanucleotide				

Repeat type	Repeat motif	Number of SSRs	Proportion in all SSRs (%)	Density (SSR/Mb)
	AAAAAT/ATTTTT	716	0.54	0.86
	AAAAAC/GTTTTT	569	0.43	0.68
	others	3178	2.38	3.8
		4463	3.34	5.33
Total	-	133727	100	159.77

3.2 SSR marker development and primer design

To ascertain the novelty of the designed SSR primer pairs, the SSR-containing sequences from the genomic sequence of sweet potato cv. Taizhong6 was analyzed. A total of 133,727 SSRs were identified for mono- to hexanucleotide repeats. Of these repeat sequences, dinucleotides (38.50%), trinucleotides (31.45%) and mononucleotides (12.77%) represented 82.71% of the total SSRs (Table 2; Additional file 2: Figure A1). Based on the SSRs obtained, a total of 2,431 SSR primer pairs were designed using Primer3 ver. 4.0.0 (http://bioinfo.ut.ee/primer3/).

3.3 SSR marker validation

The efficiency of the SSR primer pairs was evaluated and validated in 24 sweet potato materials using 100 randomly selected primer pairs. Out of these primers, 50 primer pairs (50%) effectively amplified clear bands. The amplification results showed that the 50 primer pairs amplified 251 alleles in the 24 sweet potato materials with a mean allele number of 5.02 per locus for the values ranging between 1 and 13 (Table 3). An example of the PCR amplified products in the 24 cultivars is shown in Fig. 1. Table 3 shows the characteristics of the 50 randomly selected primers used for the study.

Table 3
Characteristics of the 50 SSR primer pairs validated in 24 sweet potato materials.

No.	Primer	Repeat motif	Primer sequence (5′–3′)	Size (bp)	Chromosome	Allele No.
1	IBM 2193	(AT)16	F: TGCATGTTTGGATGTTACAGG	248	Chr 15	4
			R: CAATTACCGGAAAATTTTGGTC			
2	IBM 289	(AT)21	F: GGACTAAAATGAGCGCGAAA	210	Chr 9	1
			R:TGAAAGAAAAATTCCAACAATCAA			
3	IBM 279	(TAT)8(TGT)13(TAT)7	F: GGTCTTTGGCCAGACTATCG	280	Chr 9	1
			R: GCGGAGATCCCTTGTCATTA			
4	IBM 222	(AT)18	F: GTTCGCAGTTCCAAATTGCT	254	Chr 7	7
			R: GCATCAACAACAAAAACAAAACA			
5	IBM 1776	(TA)21	F: ATGACACTTTGGCAAATGATAGT	245	Chr 11	1
			R: AGTAAGGCAGCCCTAACCCG			
6	IBM 244	(A)24	F: GATCCCTCGAGGTGTGAAAG	250	Chr 8	7
			R: AGGATCATGCTTCACCAACC			
7	IBM 210	(A)25	F: CCTGTCCACTGGTCTAAGGC	278	Chr 7	10
			R: GCGGTCTTCATCTTCTCTGG			
8	IBM 442	(T)33	F: AATCTGTCAGGGAGTGGTGG	236	Chr 10	2
			R: AAATGCAACCCAAACAAAGC			
9	IBM 203	(T)36	F: CCATATCAATAGGCCGTGCT	231	Chr 7	7
			R: CGAACCTCAGTGAAGACGAA			
10	IBM 182	(A)78	F: GCCTTTGCTTTTCCTCTCCT	262	Chr 6	9
			R: CCGGAAACCAGCTAATCAAA			
11	IBM 1895	(CT)12	F: AATTCAATGTGGGGTCTTGC	236	Chr 12	2
			R: GCTTGATCTAACTCGGTGGC			
12	IBM 335	(A)25	F: CTTGAACAACACCTCAGGCA	199	Chr 9	2
			R: CGAGAGGAATCAGAGCCAAC			

No.	Primer	Repeat motif	Primer sequence (5'–3')	Size (bp)	Chromosome	Allele No.
13	IBM 346	(ATT)13	F: ATGCCCACATCATCATCATC	254	Chr 9	2
			R: GAATCACATATTTGCCCCTGA			
14	IBM 56	(TA)11	F: GGCCTTAGTCTTCGAAACACAT	236	Chr 2	4
			R: CGTTTGGTCTTCTGGGGTTA			
15	IBM 228	(TGA)8	F: CTCTCTTTCTTCCTTTGCCG	259	Chr 7	4
			R: GGTAGAGAAGGGAGGAGAAAGG			
16	IBM 126	(CA)12	F: GCGAAAATGTCACCGAGTTT	191	Chr 4	7
			R:GCTCTTTTCTCATCGCACCT			
17	IBM 422	(A)25	F: TTGTTCTTGCCCAATTTGCT	211	Chr 9	2
			R: AAACAATCAGCCCACACACA			
18	IBM 89	(T)26(T)30	F: CTTAGCGCTTCATGGGAGAC	259	Chr 3	4
			R: GGCATAATCAGCTCAATTCCA			
19	IBM 97	(TTAAAA)4	F: TGCAATTAGGCTACCGAACC	180	Chr 3	1
			R: GTCTCCGGTGAGACGTGTTT			
20	IBM 127	(TAA)8	F: TTCATCCTGCAAACACATGC	266	Chr 4	6
			R: TTAACGCCAACCCAACTTTC			
21	IBM 265	(T)27	F: AAACTTAGGTGATCCCAATCC	209	Chr 8	13
			R: AACATAGTTGGTTCGTCGCC			
22	IBM 185	(T)52	F: TACGTTGTCTTCCCTTCCCA	224	Chr 6	8
			R: TTGGAATTACATCAACCCCC			
23	IBM 2166	(TA)15	F: TGGGTTGAGGTTGAGGAAAC	208	Chr 15	3
			R: CTTCTAAAACCATCGCCCAA			
24	IBM 59	(AT)12	F: ATCCAATGACGCTAGTTCGG	175	Chr 2	3
			R: CCAAAAACACAGCCATCAGA			
25	IBM 247	(CTTT)6	F: TTTGGAGGCCCACTACAAAG	206	Chr 8	1

No.	Primer	Repeat motif	Primer sequence (5'–3')	Size (bp)	Chromosome	Allele No.
			R: CAGTGCATGATGGACCATTG			
26	IBM 124	(T)26(T)36	F: TCTTGAAGGGGTAAGGCAAA	234	Chr 4	4
			R: CATAAAATTTTGCTCCACATGC			
27	IBM 2060	(CAA)10(ATA)16	F: AAGAAATCTTTTTGGAATGCGA	255	Chr 14	4
			R: ACCGTACAACGACGGTTCAT			
28	IBM 1745	(ATAC)6(AT)16	F: TGTGTTTGGTTCAACAAGGAA	244	Chr 11	7
			R: ACGAGTTGGGTATGAATCGG			
29	IBM 2209	(TTG)8	F: ATGGTTTTGTGGGCAAAGT	138	Chr 15	1
			R: ACGCTCTCTTCATGCCAAGT			
30	IBM 2010	(TAT)14(TGTTAT)6	F: TTAATAAAAGTTTGCGCGGG	208	Chr 13	7
			R: ATGCAGATCCCTGATTTTGG			
31	IBM 255	(TAT)16	F:AAATTTATTTAGATTTTGGATACGGA	234	Chr 8	5
			R: ATTGTTACCATGCACAGGCA			
32	IBM 211	(A)23(A)21	F: GACACTGAATTGATCTCCCGA	207	Chr 7	10
			R: TCGGTTGTTGTTGTTGTTTTT			
33	IBM 760	(T)25	F: GCCAGAATTTTCTGTCAACCA	180	Chr 10	1
			R: AAAAGAACGTGGGGAAGGAA			
34	IBM 1174	(A)25	F: TGCAACATGCCATAAATGCT	267	Chr 10	2
			R: CCTAAAGCTTTCCCGTTTTG			
35	IBM 106	(T)87	F: TTGGGGAAGGCTTTTAGGTT	280	Chr 3	10
			R: TTGTGATCCTTTCTCAGTTAAGGT			
36	IBM 52	(TTA)11	F: GCACTTAGCCACCCCTATCA	171	Chr 2	10
			R: AAACAAAATTGTGGGAGAGCA			
37	IBM 261	(TAA)9(TAT)8 (ATC)7	F: TGCATTTAAAAACTCCGTAATACA	218	Chr 8	9
			R: GAATGAATGCAATTCTAAAAACCC			

No.	Primer	Repeat motif	Primer sequence (5'–3')	Size (bp)	Chromosome	Allele No.
38	IBM 27	(A)24	F: GGTTTGAATTTGGAGTGAACATC	217	Chr 1	1
			R: TGAGTTGTGACGTGTGAGCA			
39	IBM 2082	(AT)16	F: TATCTACCCAACCGACCTGC	238	Chr 14	1
			R: CCGTTAGATCTGAACACGTGAA			
40	IBM 241	(T)25	F: CTGCACACATGCAACACAAC	189	Chr 8	10
			R: TCAGTATCACAAAGCTCCACAA			
41	IBM 1923	(ATT)9	F: CAACCAAACCCCCTAATGTG	211	Chr 13	4
			R: ACATGGTTTCAGAGGGACCA			
42	IBM 1733	(TA)19	F: TGATTTTGGATGTTATTTCATCATTT	270	Chr 11	6
			R: TCTTGGCTTAAGTTATCGGCA			
43	IBM 119	(A)24	F:GGAAACGTTAGTACAAGTTGACACA	273	Chr 4	4
			R: TCGCACATTATTAAAAACGGTCT			
44	IBM 1895	(CT)12	F: AATTCAATGTGGGGTCTTGC	236	Chr 12	2
			R: GCTTGATCTAACTCGGTGGC			
45	IBM 291	(A)30	F: CCAAGCAAGCACACAACTTT	277	Chr 9	7
			R: GCACGCTGTGCTTAAAATGA			
46	IBM 53	(A)45	F: CCAAACACCCACATAGACACC	239	Chr 2	10
			R: AAGCACACTGATGTGCCACT			
47	IBM 582	(ATT)9	F: AAGGTTATGATGGCCGACTG	157	Chr 10	7
			R: AAAAACTCCGTTCCCATCAA			
48	IBM 296	(A)25	F: ATAAGAAGAGAGCGGGTCGG	220	Chr 9	5
			R: TGCACTTTGAATGCACAACA			
49	IBM 1984	(AT)23	F: TGACATGTGCCGATACTCTAAAA	250	Chr 13	3
			R: GCAAAACACTTCTTCATGGG			
50	IBM 32	(T)29	F: TCCACATAAGGGAGATGAGGA	252	Chr 1	10

No.	Primer	Repeat motif	Primer sequence (5′–3′)	Size (bp)	Chromosome	Allele No.
			R: TGTGGAGGGGAGAGAGTGTT			
Total						251
Mean						5.02

3.4 Cluster analysis

The genetic diversity of 24 sweet potato varieties as analyzed by the 50 primer pairs revealed a genetic distance range of 0.605 to 1.00 with an average distance of 0.740 among the 24 sweet potato varieties studied (Additional file 1: Table A1). According to the UPGMA clustering results, the genetic similarity coefficient between the 24 sweet potato germplasms was relatively high ranging between 0.66 and 0.87 with an average of 0.765 (Fig. 2). The dendrogram generated grouped the 24 resources into four (4) clusters (cluster I-IV; Fig. 2) at a similarity coefficient of 0.68. Among them, dayehong and fushu18 were the same with the closest genetic relationship and the similarity coefficient was 0.745. The clustering results revealed no direct relationship with the geographical sources of germplasm, indicating a more frequent exchange of germplasm in sweet potato cultivation and breeding.

4.0 Discussion

The advancement of next-generation sequencing has generated scores of datasets for many plant species that provide useful genomic materials for developing efficient molecular markers for genetic analyses [27]. Recently, technological advancement in high throughput DNA sequencing offers new information to accelerate the development of molecular markers. SSR markers are more efficient in constructing linkage maps and for genetic analyses in crops owing to their high polymorphism and specificity [27]. Many SSR markers have previously been designed and used for variety identification, genetic diversity analysis and construction of linkage maps in sweet potato [28–30] compared to other molecular markers as a result of its co-dominant nature, consistency, high polymorphism, and reproducibility [31]. Although several studies have reported the use of SSR markers in sweet potato, their number and availability are limited with only a few being polymorphic compared to other crops.

To identify valuable SSR markers for sweet potato genetic improvement, the sweet potato genome was searched and a total of 2,431 SSR markers were successfully developed. The distribution density was 159.77 Mb per SSR or 6.26 kb per one SSR on average which was lower than the average density recorded for pigeon pea (8.4 kb), cotton (20.0 kb), and soybean (23.80 kb) but almost the same as that of sesame (6.55 kb), and relatively higher compared to that of rice (3.4 kb) and radish (4.93 kb) [32]. However, the differences in frequency and abundance could be attributed to the size of the database, tools for SSR data-mining, the length of repeat motifs and the application of different repeat unit thresholds [33]. In our current study, mono-, di- and trinucleotides were the most common SSRs with dinucleotides showing the highest frequency (38.50%) followed by trinucleotides (31.46%) and mononucleotide (12.77%; Additional file 2: Figure A1). This finding contrasts with previous reports showing trinucleotides as the most dominant repeat motifs in sweet potato followed by dinucleotides [10, 13]. Other studies also suggested trinucleotides as the second predominant repeat motifs in sweet potato which is in agreement with our current findings [12].

The main repeat types among the identified SSRs were A/T (12.61%), AT/AT (61.51%), AAT/ATT (27.42%), and AAAT/ATTT (13.32%, Additional 2: Figure A1). In agreement with our current study, Wang et al. [10] identified AAT/ATT as the dominant SSR motif in the sweet potato. Similarly, Yang et al. [27] identified AAAT/ATTT as the most frequent repeat motif among tetranucleotides in Welsh onion. However, previous studies identified AG/CT, AAG/CTT, and AT/TA motifs as the most dominant motif types in sweet potato [34], conflicting with our findings.

Out of the 100 randomly selected primer pairs screened, 50 primer pairs (50%) produced clear stable bands and amplified 251 alleles in the 24 sweet potato materials (Table 4). The number of alleles recorded was 5.02 per locus on average for values ranging between 1 and 13. Consistent with our results, previous studies reported a higher number of alleles per locus using SSR markers to analyze the genetic diversity of sweet potato germplasm. This indicates a high polymorphism among the sweet potato accessions studied. Yada et al. [35] had 2 to 6 alleles per primer while Buteler et al. [36] recorded 3 to 10 alleles per SSR and high polymorphism in sweet potato. Roullier et al. [14] reported 4 to 23 alleles per locus among 329 accessions from South America, Central America, and the Caribbean using 13 SSR primers; Veasey et al. [37] also reported 3 to 10 alleles per locus in Brazilian sweet potato accessions while Tumwegamire et al. [15] reported 2 to 11 per locus in East African sweet potato varieties. Conversely, Hwang et al. [38] had low polymorphism and recorded 1 to 4 alleles per SSR using varied annealing temperatures and SSR primers. The result of our current study confirms the exceptional discriminatory ability of SSR markers [39].

As a hexaploid plant distinguishing between homozygous and heterozygous sites becomes difficult hence dominant markers are preferred over collinear markers [8, 9]. Previous studies reported the high polymorphism of sweet potato which is attributed to the large genome size and high heterozygosity [38] influenced by its mating systems (self-incompatibility and outcrossing). Again, the polyploidy (autohexaploid) of sweet potato combined with the large chromosome number (2n = 6x = 90) makes sweet potato SSR primers highly polymorphic [30, 40]. Hence, it is likely for sweet potato genotypes to have huge genetic distances among them even in smaller populations [41]. Overall, vegetative propagation could be very effective in conserving the genetic diversity and heterozygosity of sweet potato.

The average SSR-based genetic distance among the 24 sweet potato varieties was 0.740 on average for values ranging between 0.605 and 1.00 (Additional file 1: Table A1). The genetic similarity coefficient range of 0.66 to 0.87 with a mean value of 0.765 recorded in this study is high, indicating a low diversity in the sweet potato materials studied (Fig. 2). The result is consistent with Hwang et al. [38] who recorded a high similarity coefficient of 0.64 on average and thus, concluded a low diversity among the accessions studied. On the contrary, Yada et al. [35] reported an average similarity coefficient of 0.57 by evaluating the genetic diversity of cultivars from Uganda. Zhang et al. [42] observed a low similarity coefficient (0.588) amid sweet potato varieties from South America. Tumwegamire et al. [15] also recorded a similarity coefficient of 0.54 on average when the genetic diversity of farmer varieties of both white- and orange-fleshed sweet potato from East Africa were assessed. Similarly, David et al. [43] reported a low genetic similarity coefficient of 0.54 on average and concluded a high diversity among the studied accessions. Thus, the differences could be attributed to the number and type of markers used and the genotypic variances. The clustering results revealed no direct relationship between the national and regional sources of germplasm, indicating a more frequent exchange of germplasm in sweet potato cultivation and breeding. The above results show the effectiveness and feasibility of the developed SSR markers for assessing genetic diversity in sweet potato.

5.0 Conclusions

To facilitate marker-assisted selection (MAS) and explore new molecular markers in sweet potato, we developed a set of SSR markers from the reference genome of cultivated sweet potato (cv. Taizhong6) using MISA software. A total of 133,727 SSRs were identified and 2,431 new SSR markers were developed. 50% of the randomly selected showed good amplification effects and produced clear stable bands. The ability of the markers to identify specific traits in different genotypes and analogize cultivars make them a valuable resource for germplasm identification, genetic relationship studies, and diversity analysis in sweet potato and other related species. Moreover, the markers are a significant addition to the currently available SSR markers for sweet potato genetic studies.

Declarations

Author Contributions

HAD conceptualized the idea, led the experimental work, developed and screened SSR markers, designed primers, analyzed genotypic data and drafted the manuscript. JZ and CS collected experimental materials, assisted with the experimental work and statistical analysis. ZH proofread and revised the manuscript. All authors read and approved the final version of the manuscript.

Funding

This research was funded by the National Natural Science Foundation of China-Guangdong Joint Fund, China, and Studies on Resistance Resources and Molecular Mechanisms of Sweet potato Weevil in South China **(Grant number, U1701234)**.

Acknowledgments

We are thankful to Mr. Eric Amenyogbe and Mr. Gyan Ray Watson for critical review and useful suggestions during the manuscript preparation.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Rozen S, Skaletsky H: **Primer3 on the WWW for general users and for biologist programmers**. In: *Bioinformatics methods and protocols*. Springer; 2000: 365–386.

2. Lebot V, Michalet S, Legendre L. Identification and quantification of phenolic compounds responsible for the antioxidant activity of sweet potatoes with different flesh colours using high performance thin layer chromatography (HPTLC). J Food Compos Anal. 2016;49:94–101.

3. Chandrasekara A, Josheph Kumar T: **Roots and tuber crops as functional foods: a review on phytochemical constituents and their potential health benefits**. *International journal of food science* 2016, **2016**.

4. Yan L, Lai X, Li X, Wei C, Tan X, Zhang Y. Analyses of the complete genome and gene expression of chloroplast of sweet potato [Ipomoea batata]. PLoS One. 2015;10(4):e0124083.

5. Kumar P, Gupta V, Misra A, Modi D, Pandey B. Potential of molecular markers in plant biotechnology. Plant Omics. 2009;2(4):141.

6. Kriegner A, Cervantes JC, Burg K, Mwanga RO, Zhang D. A genetic linkage map of sweetpotato [Ipomoea batatas (L.) Lam.] based on AFLP markers. Mol Breeding. 2003;11(3):169–85.

7. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA: **The sequence of the human genome**. *science* 2001, **291**(5507):1304–1351.

8. Yue X-y, Liu G-q, Zong Y, Teng Y-w. Cai D-y: **Development of genic SSR markers from transcriptome sequencing of pear buds**. Journal of Zhejiang University Science B. 2014;15(4):303–12.

9. Silva PI, Martins AM, Gouvea EG, Pessoa-Filho M, Ferreira ME. Development and validation of microsatellite markers for Brachiaria ruziziensis obtained by partial genome assembly of Illumina single-end reads. Bmc Genomics. 2013;14(1):17.

10. Wang Z, Li J, Luo Z, Huang L, Chen X, Fang B, Li Y, Chen J, Zhang X: **Characterization and development of EST-derived SSR markers in cultivated sweetpotato (Ipomoea batatas)**. *BMC plant biology* 2011, **11**(1):139.

11. Hu J, Nakatani M, Mizuno K, Fujimura T. Development and characterization of microsatellite markers in sweetpotato. Breed Sci. 2004;54(2):177–88.

12. Xie F, Burklew CE, Yang Y, Liu M, Xiao P, Zhang B, Qiu D. De novo sequencing and a comprehensive analysis of purple sweet potato (Impomoea batatas L.) transcriptome. Planta. 2012;236(1):101–13.

13. Tao X, Gu Y-H, Wang H-Y, Zheng W, Li X, Zhao C-W, Zhang Y-Z. Digital gene expression analysis based on integrated de novo transcriptome assembly of sweet potato [Ipomoea batatas (L.) Lam.]. PloS one. 2012;7(4):e36234.

14. Roullier C, Rossel G, Tay D, McKey D, Lebot V. Combining chloroplast and nuclear microsatellites to investigate origin and dispersal of New World sweet potato landraces. Mol Ecol. 2011;20(19):3963–77.

15. Tumwegamire S, Rubaihayo P, LaBonte D, Diaz F, Kapinga R, Mwanga R, Grüneberg W. Genetic diversity in white-and orange-fleshed sweetpotato farmer varieties from East Africa evaluated by simple sequence repeat markers. Crop Sci. 2011;51(3):1132–42.

16. Zhao N, Yu X, Jie Q, Li H, Li H, Hu J, Zhai H, He S, Liu Q. A genetic linkage map based on AFLP and SSR markers and mapping of QTL for dry-matter content in sweetpotato. Molecular breeding. 2013;32(4):807–20.

17. Schafleitner R, Tincopa LR, Palomino O, Rossel G, Robles RF, Alagon R, Rivera C, Quispe C, Rojas L, Pacheco JA. A sweetpotato gene index established by de novo assembly of pyrosequencing and Sanger sequences and mining for gene-based microsatellite markers. BMC Genomics. 2010;11(1):604.

18. Porebski S, Bailey LG, Baum BR. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant molecular biology reporter. 1997;15(1):8–15.

19. Thiel T, Michalek W, Varshney R, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). Theoretical applied genetics. 2003;106(3):411–22.

20. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. Primer3—new capabilities and interfaces. Nucleic acids research. 2012;40(15):e115–5.

21. Xie J, Zhang M, Sun J, Zhang F. Genome-wide genic SSR marker development for the endangered Dongxiang wild rice (Oryza rufipogon). Plant Genetic Resources. 2017;15(6):566–9.

22. Han YC, Teng CZ, Hu ZL, Song YC. An optimal method of DNA silver staining in polyacrylamide gels. Electrophoresis. 2008;29(6):1355–8.

23. Chevallet M, Luche S, Rabilloud T. Silver staining of proteins in polyacrylamide gels. Nature protocols. 2006;1(4):1852.

24. Liu K, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics. 2005;21(9):2128–9.

25. Nei M, Li W-H: Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences* 1979, **76**(10):5269–5273.

26. Rohlf F: NTSYS 2.1: numerical taxonomic and multivariate analysis system. *New York, Exeter Software* 2000.

27. Yang L, Wen C, Zhao H, Liu Q, Yang J, Liu L, Wang Y. Development of polymorphic genic SSR markers by transcriptome sequencing in the welsh onion (Allium fistulosum L.). Applied Sciences. 2015;5(4):1050–63.

28. Yada B, Brown-Guedira G, Alajo A, Ssemakula GN, Mwanga RO, Yencho GC. Simple sequence repeat marker analysis of genetic diversity among progeny of a biparental mapping population of sweetpotato. HortScience. 2015;50(8):1143–7.

29. Yang X, Su W, Wang L, Lei J, Chai S, Liu Q. Molecular diversity and genetic structure of 380 sweetpotato accessions as revealed by SSR markers. J Integr Agric. 2015;14(4):633–41.

30. Ngailo S, Shimelis H, Sibiya J, Amelework B, Mtunda K. Genetic diversity assessment of Tanzanian sweetpotato genotypes using simple sequence repeat markers. South African Journal of Botany. 2016;102:40–5.

31. Karihaloo J: **DNA fingerprinting techniques for plant identification**. In: *Plant Biology and Biotechnology*. Springer; 2015: 205–221.

32. Zhai L, Xu L, Wang Y, Cheng H, Chen Y, Gong Y, Liu L. Novel and useful genic-SSR markers from de novo transcriptome sequencing of radish (Raphanus sativus L.). Mol Breeding. 2014;33(3):611–24.

33. Dutta S, Kumawat G, Singh BP, Gupta DK, Singh S, Dogra V, Gaikwad K, Sharma TR, Raje RS, Bandhopadhya TK. Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [Cajanus cajan (L.) Millspaugh]. BMC plant biology. 2011;11(1):17.

34. Zhang K, Wu Z, Tang D, Lv C, Luo K, Zhao Y, Liu X, Huang Y, Wang J. Development and identification of SSR markers associated with starch properties and β-carotene content in the storage root of sweet potato (Ipomoea batatas L.). Frontiers in plant science. 2016;7:223.

35. Yada B, Tukamuhabwa P, Wanjala B, Kim D-J, Skilton RA, Alajo A, Mwanga RO. Characterization of Ugandan sweetpotato germplasm using fluorescent labeled simple sequence repeat markers. HortScience. 2010;45(2):225–30.

36. Buteler M, Jarret R, LaBonte D. Sequence characterization of microsatellites in diploid and polyploid Ipomoea. Theor Appl Genet. 1999;99(1–2):123–32.

37. Veasey EA, Borges A, Rosa MS, Queiroz-Silva JR, Bressan EdA, Peroni N. Genetic diversity in Brazilian sweet potato (Ipomoea batatas (L.) Lam., Solanales, Convolvulaceae) landraces assessed with microsatellite markers. Genetics Molecular Biology. 2008;31(3):725–33.

38. Hwang SY, Tseng YT, Lo HF. Application of simple sequence repeats in determining the genetic relationships of cultivars used in sweet potato polycross breeding in Taiwan. Sci Hortic. 2002;93(3–4):215–24.

39. Gichuru V, Aritua V, Lubega G, Edema R, Adipala E, Rubaihayo P: **A preliminary analysis of diversity among East African sweetpotato landraces using morphological and simple sequence repeats (SSR) markers**. In: *II International Symposium on Sweetpotato and Cassava: Innovative Technologies for Commercialization 703*: 2005; 2005: 159–164.

40. Li R, Zhai H, Kang C, Liu D, He S, Liu Q: **De novo transcriptome sequencing of the orange-fleshed sweet potato and analysis of differentially expressed genes related to carotenoid biosynthesis**. *International journal of genomics* 2015, **2015**.

41. Gruneberg W, Ma D, Mwanga R, Carey E, Huamani K, Diaz F, Eyzaguirre R, Guaf E, Jusuf M, Karuniawan A: **Advances in sweetpotato breeding from** 1992 **to 2012**. In.: CABI International; 2015.

42. Zhang D, Carbajulca D, Ojeda L, Rossel G, Milla S, Herrera C, Ghislain M: **Microsatellite analysis of genetic diversity in sweetpotato varieties from Latin America**. *CIP Program report* 2000, **2000**(2000):295–301.

43. David MC, Diaz FC, Mwanga RO, Tumwegamire S, Mansilla RC, Grüneberg WJ. Gene pool subdivision of East African sweetpotato parental material. Crop Sci. 2018;58(6):2302–14.

Additional Files

Additional file 1

Table A1: Genetic distances generated from the SSR marker analysis of sweet potato germplasm (1-24) in this study.

Additional file 2

Figure A1: Characteristics of SSRs in the sweet potato genome.

   a. Distribution of different repeat types

b. Frequency distribution of major repeat motifs

c. Frequency distribution of main motif sequence in **I.** Dinucleotide, **II.** Trinucleotide and **III.** Tetranucleotide repeats

Figures
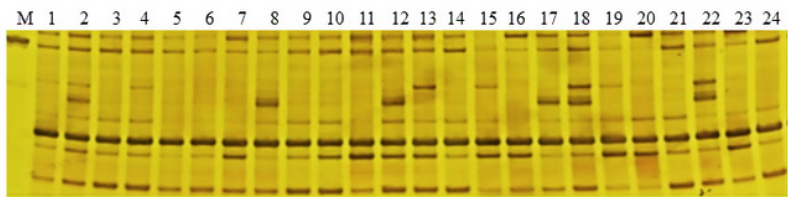
Figure 1

Amplification results generated by IBM 241 in the 24 sweet potato varieties. The numbers 1–24 represent sweet potato materials and M represents the DNA marker.
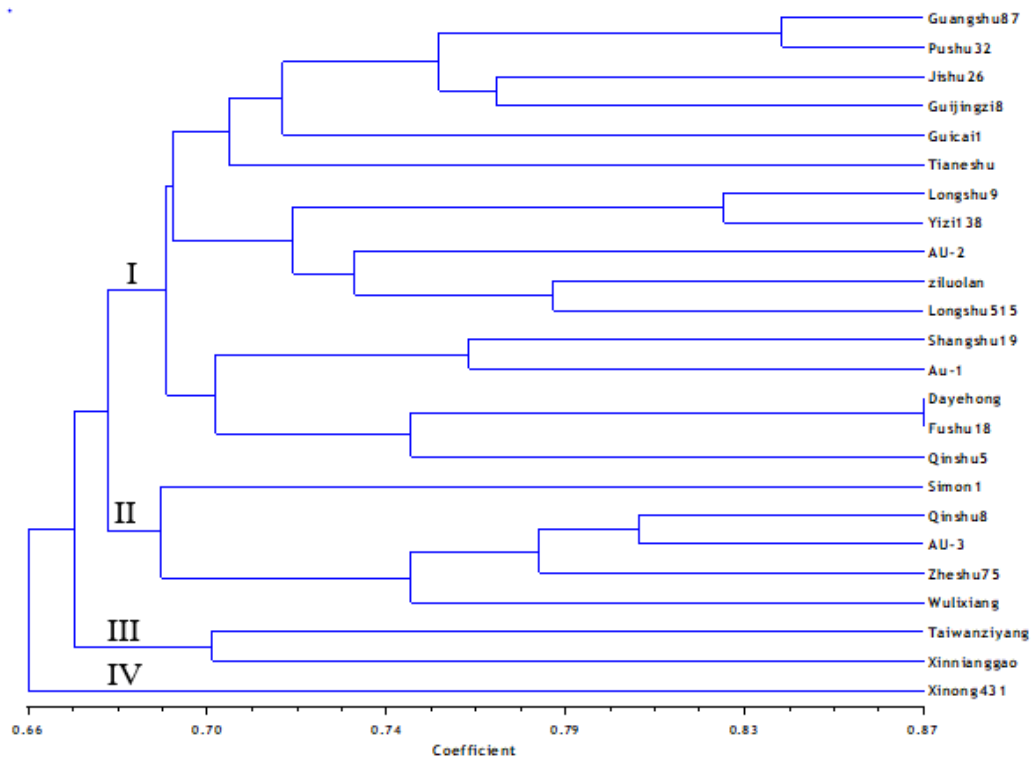
Figure 2

UPGMA dendrogram of the genetic relationships among 24 sweet potato accessions.