

Article

Development of Simple Sequence Repeat (SSR) Markers of Sesame (*Sesamum indicum*) from a Genome Survey

Xin Wei †, Linhai Wang †, Yanxin Zhang, Xiaoqiong Qi, Xiaoling Wang, Xia Ding, Jing Zhang and Xiurong Zhang *

Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Key Laboratory of Biology and Genetic Improvement of Oil Crops of the Ministry of Agriculture, Wuhan 430062, China

† These authors contributed equally to this work.

* Author to whom correspondence should be addressed; E-Mail: Zhangxr@oilcrops.cn; Tel./Fax: +86-27-86711836.

Received: 17 February 2014; in revised form: 31 March 2014 / Accepted: 14 April 2014 /

Published: 22 April 2014

Abstract: Sesame (*Sesamum indicum*), an important oil crop, is widely grown in tropical and subtropical regions. It provides part of the daily edible oil allowance for almost half of the world's population. A limited number of co-dominant markers has been developed and applied in sesame genetic diversity and germplasm identity studies. Here we report for the first time a whole genome survey used to develop simple sequence repeat (SSR) markers and to detect the genetic diversity of sesame germplasm. From the initial assembled sesame genome, 23,438 SSRs (≥ 5 repeats) were identified. The most common repeat motif was dinucleotide with a frequency of 84.24%, followed by 13.53% trinucleotide, 1.65% tetranucleotide, 0.3% pentanucleotide and 0.28% hexanucleotide motifs. From 1500 designed and synthesised primer pairs, 218 polymorphic SSRs were developed and used to screen 31 sesame accessions that from 12 countries. STRUCTURE and phylogenetic analyses indicated that all sesame accessions could be divided into two groups: one mainly from China and another from other countries. Cluster analysis classified Chinese major sesame varieties into three groups. These novel SSR markers are a useful tool for genetic linkage map construction, genetic diversity detection, and marker-assisted selective sesame breeding.

Keywords: SSR; genetic diversity; genome survey; sesame

1. Introduction

Sesame (*Sesamum indicum* L.) is one of the most commonly grown oil crops and has been suggested as the most ancient oil crop [1]. It provides part of the daily oil and protein needs for almost half of the world's population. It is also directly consumed directly as a leafy vegetable, sweetmeat, cooking ingredient and snack food. Its high regard by consumers has earned it the poetic label “queen of oilseeds” [2] for its high oil content, high oil quality, and strong resistance to oxidation. Sesame oil content in sesame (up to 64% of dry seed) is much higher than that in other edible oil crops, such as soybean (~20%), rapeseed (~40%) and maize (~8%) [3]. The amount of unsaturated fatty acid in sesame oil (>80%), which is beneficial to human health, is also greater than that found in other main oil crops [4]. Moreover, some strong antioxidant compositions (e.g., sesamol, lignan, sesamin, and sesamolol) are rich in sesame oil and possess many potent pharmacological properties leading to physiological effects such as decreased blood lipids and lowered cholesterol levels [5–7].

Several types of molecular markers have been developed and applied to sesame genotyping, such as random amplified polymorphic DNA (RAPD) [8,9], inter-simple sequence repeats (ISSR) [10], amplified fragment length polymorphism (AFLP) [11], sequence-related amplified polymorphisms (SRAP) [12] and expressed sequence tags-SSR (EST-SSR) [13]. However, most studies were carried out using random molecular markers and few co-dominant polymorphic markers were used.

SSR markers are widely used for high-throughput genotyping and map construction as they are advantageous due to high abundance, random distribution within the genome, high polymorphism information content (PIC) and stable co-dominance [14–16]. The reproducibility, co-dominance, relative abundance and complete genome coverage of SSR markers have made them one of the most useful tools for detecting genetic diversity, genetic linkage mapping, association mapping and evolution analysis [17]. Genomic SSRs and expressed sequence tag (EST)-SSRs, which are considered complementary to plant genome mapping, have been reported in several primary oil crops, such as rapeseed [18], peanut [19] and soybean [20]. Few EST-SSRs were developed and used to detect genetic diversity for sesame germplasm [13,21,22]. However, their use is limited due to relatively low polymorphism and high possibility of no gene-rich regions in the genome. In contrast, genomic SSRs are highly polymorphic and tend to be widely distributed throughout the genome, resulting in more accurate detection of genetic diversity [23].

An elite cultivar, Zhongzhi13, which was introduced to most major sesame planting areas in China over the last 10 years, was sequenced *de novo* by our group using the Illumina Hiseq2000 platform [24]. After reads filtering, 54.5 Gb of high-quality data were obtained, representing approximately 152.7 folds of the sesame genome. The Zhongzhi 13 draft genome was 274 Mb, covering 76.7% of the genome size, which was derived from a 17-mer depth distribution (357 Mb). In this study, sesame genomic SSRs of sesame were identified and developed by whole genome survey. We mined and validated 218 SSR markers and described their application for understanding the genetic relationship of 31 sesame accessions from worldwide. Thirty-two high polymorphism SSR markers were selected as the core SSR markers and used to detect the relationship among 23 major Chinese sesame varieties. These novel polymorphism SSR markers were effectively applied to genotype and genetic diversity analyses.

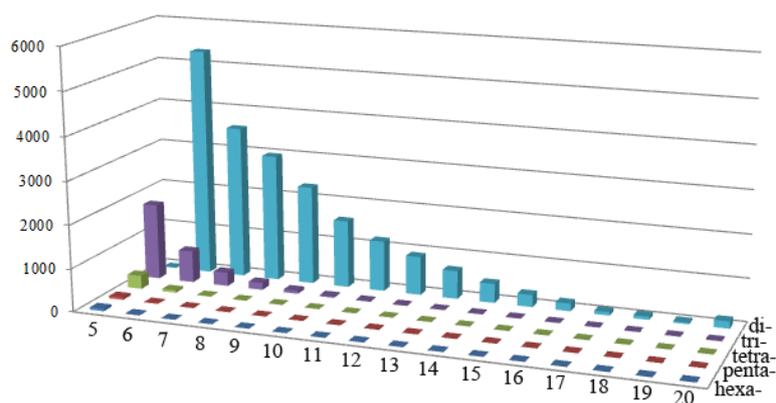
2. Results and Discussion

2.1. Frequency Distribution of Different SSR Markers Types

The assembled scaffolds were searched for the presence of SSRs in the genome using the MISA software, and 23,438 putative SSRs from 196 of 1,036 scaffolds (≥ 2 kb) were identified. The average distance between SSRs was approximately 11.69 kb in sesame genomic DNA. SSRs in plant genomes have been surveyed in many species, such as *Oryza sativa*, *Arabidopsis thaliana* and *Sorghum bicolor* [25]. Genomic SSR number and frequency were quite different among these sequenced plants. SSR numbers in *O. sativa*, *A. thaliana*, *S. bicolor*, *S. indicum* were 70,531, 15,249, 73,658 and 23,438, respectively; the frequencies of SSR occurrence in these species in each locus were 2.75 kb, 2.39 kb, 5.70 kb and 11.69 kb, respectively. SSR frequency in *A. thaliana* was nearly five times greater than that in *S. indicum*.

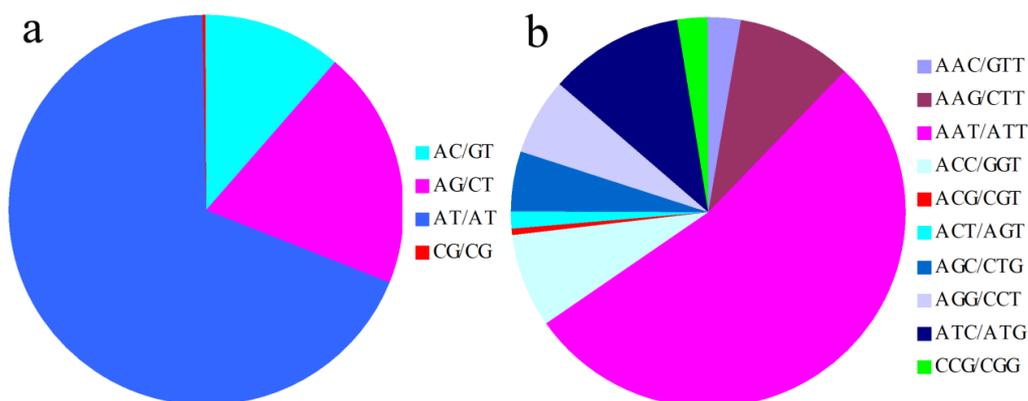
SSR loci contained in scaffolds ranged from 1 to 1,049 (average of 119), and almost all scaffolds (192) contained more than one SSR. Among the derived SSR repeats, the dinucleotide motif was present in all scaffolds: trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide motifs existed in 155 (79.08%), 110 (56.12%), 48 (24.49%), and 42 (21.43%) scaffolds, respectively. Short tandem repeats categorised by their unit sizes and the number of repeats are summarized in Figure 1. Dinucleotide repeats (84.24%) comprise the largest group of repeat motifs, accounting for more than three-quarters of the total SSR content. Less commonly found repeats were trinucleotide repeats (13.53%), followed by tetranucleotide repeats (1.65%) and pentanucleotide repeats (0.30%), with hexanucleotide repeats being the least abundant at merely 0.28% (Figure 1). A large proportion of both dinucleotides and trinucleotides was observed, while the rest amounted to less than 2.3%. Although the frequency was diverse, the overall pattern of SSR motifs of particular lengths was similar. Dinucleotide repeats dominated over other types of repeats in all the species and SSR frequency decreased stepwise with increased motif length. Dinucleotide repeats were found to be more than 50% in all genomes. While the dinucleotide and trinucleotide repeats mostly contributed to the major proportion of SSRs, a very small share was contributed by tetranucleotide, pentanucleotide and hexanucleotide repeats. Because high throughput sequencing mainly acquires short read lengths during sequencing, the presence of repeats with large unit lengths might have been decreased.

Figure 1. Distribution of various classes of simple repeat motifs with different numbers of repeats in the sesame genome.



SSR motifs consisted of 102 types. Among the dinucleotide repeat motifs, the AT repeat was the most common, representing 68.80%, followed by 19.60% AG/CT repeats (Figure 2a). The predominant trinucleotide motifs (AAT/ATT, ATG/ATC and AAG/CTT) repeats accounted for 59.22%, 11.31% and 9.36%, respectively (Figure 2b). Among the dinucleotide repeat motifs, (AT)_n, which was up to 68.80%, was the most common repeat unit in *S. indicum*, (AT)_n is also the most common repeat in *A. thaliana*, *S. bicolor*, and *Brassica napus* [25,26].

Figure 2. (a) Percentage of different dinucleotide repeat motifs in the sesame genome. and (b) Percentage of different motifs in trinucleotide repeat motifs in the sesame genome.



2.2. SSR Marker Polymorphism

We first designed and synthesized 1,500 SSR primer pairs from those scaffolds greater than 2 kb. The majority of SSR loci were dinucleotide repeats (1,097, 73.13%), and trinucleotide repeats (255, 17.00%). Initially, four representative cultivars (Aizhima, Mishuozhima, Gaozhima and Silengzhima) were selected to detect the effectiveness of these primer pairs in through denaturing polyacrylamide gel electrophoresis (PAGE). The four landraces included major sesame characteristics, such as ramiform, rhabdus, black, white, dwarf and tall individuals. Finally, 1,449 (96.60%) SSR loci were amplified successfully in all four accessions, and 218 (14.53%) of them were identified as polymorphic loci in the four germplasm resources.

The frequency of each motif of the 218 polymorphic SSRs is presented in Table S1. Only 13 motif types were included. Most of them were dinucleotide repeats (87.61%), and trinucleotide, tetranucleotide, pentanucleotide, hexanucleotide repeats comprised 9.17%, 2.29%, 0.46% and 0.46%, respectively. Dinucleotide loci in the polymorphism group were greater than that in the genome, indicating that dinucleotide loci had higher diversity than other loci. Among the dinucleotide and trinucleotide repeat motifs, the AT and AAT/ATT repeats, which were up to 79.58% and 65.00%, respectively, were still the most common types.

Subsequently, 31 worldwide accessions of sesame, including 27 landraces and four elites, were used to detect SSR loci polymorphisms. Among these accessions, those from China and abroad were 16 and 15, respectively. The number of alleles (N_A) per locus varied widely among the markers (Table S2) and ranged from 2 to 13, with an average of 5.46 alleles. The frequency of major alleles (M_{AF}) per locus was 0.1290–0.9677, with an average of 0.4305. In addition, the H_O values were 0.0000–1.0000 with an average of 0.2496, and the H_E values were 0.0635–0.9138 with an average of

0.6853. Lastly, polymorphic index content values were 0.0605–0.8901, with an average of 0.6315. The polymerase chain reaction (PCR) product size ranged from 106 to 280 bp. All the primers produced amplicons in agreement with the expected sizes. Most of the SSR primers (216 primer pairs) showed significantly deviated from Hardy-Weinberg equilibrium (HWE: $p < 0.05$).

Since traditional massive-scale DNA sequencing is relatively high-cost, low-throughput and time-consuming, the development of SSR markers was not fast and limited SSR markers have been developed. Based on mass sequence sesame data obtained by RNA-Seq, 40, 32 and 59 polymorphic EST-SSRs were successfully developed, respectively [21,22,27]. Considering three recently published sets of EST-SSR markers derived from transcriptome studies, available polymorphic SSR and EST-SSR markers are still less than 150. Although EST-SSRs are useful for genetic analysis, their disadvantages of relatively low polymorphisms and high concentration in gene-rich regions of the genome may limit their use [28]. At present, using high-throughput next-generation sequencing, the whole sesame genome was sequenced and SSRs in the genome were detected. For the first time, 1,500 SSR primer pairs were designed from the sesame genome survey and 1,449 (96.60%) primer pairs were effectively amplified. Furthermore, 218 polymorphic SSRs were developed and characterized, of which 137 markers with a PIC value of >0.60 were highly informative.

2.3. Genetic Relationship Analysis

Based on the 218 polymorphic SSRs, we constructed a STRUCTURE analysis to classify the 31 sesame accessions (Figure 3). As K values increased from 1 to 10, the LnP (D) value directly increased; meanwhile, the ΔK decreased at $K = 2$ (Figure S1). Thus, STRUCTURE analysis divided all accessions into two clusters. The first group, which is shown in red, included 16 accessions, with five from China, eight from East and South Asia and three from Africa. The second group included 15 accessions, 12 of which were from China and the other three accessions were all from the United States, which were induced mutations from some Chinese sesame germplasm (Table S3). Thus, the 31 sesame accessions could be divided into two groups, and Chinese sesame might have a distant relationship with that from other countries. Phylogenetic trees were constructed for the 31 sesame accessions by neighbor-joining (NJ) analysis (Figure 4) and unweighted pair group method with arithmetic mean (UPGMA) analysis (Figure S2). These sesame samples were also divided into two groups: one was mainly from China and the other was from abroad. The result was in line with the conclusion of STRUCTURE analysis.

Figure 3. Population structure of sesame accessions. STRUCTURE was constructed by all loci. $K = 2$. Clusters are indicated by different colours. Samples included in all clusters are listed in Table S3.

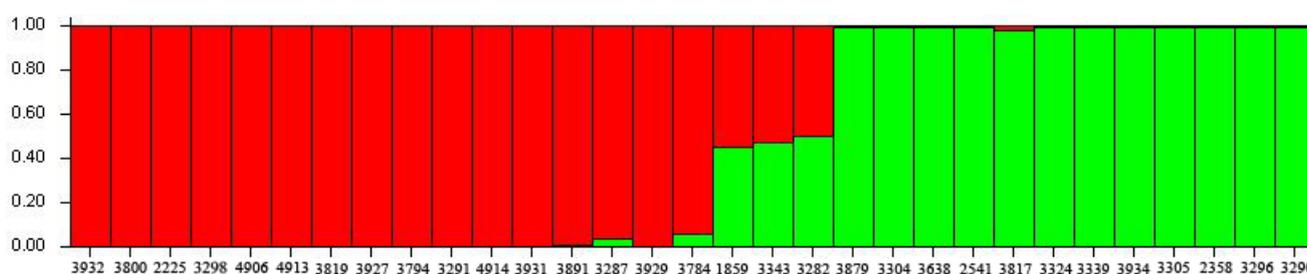
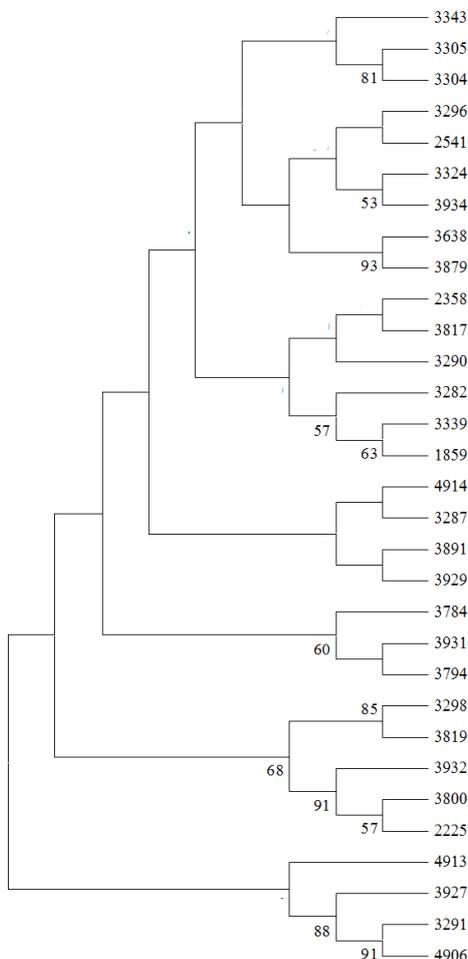


Figure 4. Phylogenetic tree of 31 sesame accessions derived from NJ analysis based on 218 SSR markers. Bootstrap values >50 are shown.



Some samples from China and other countries had not been strictly separated in the STRUCTURE and phylogenetic analyses, which may have been due to sample out-crossing. Sesame is mainly self-pollinated, although outcrossing between 5% and 60% has been reported [29]. The sesame samples from other countries and from China were planted together and a few sesame samples might have crossed with each other. However, most samples from China and other countries were still clustered into different groups in the analysis.

Linkage disequilibrium (LD) is the correlation between polymorphisms that is caused by their shared history of mutation and recombination. It plays a central role in association analysis [30]. The allele frequency correlation (r^2) between segregation sites was used to measure the degree of LD. The level of r^2 for all sesame SSR loci was shown in Figure S3. The extent of LD was low for most SSR loci, indicating that LD between SSRs in sesame was weak. Among the 23653 loci pairs, only 1006 (4.25%) of loci pairs were in significant LD at the comparison-wise 0.05 level. This value was less than that in maize [31]. What's more, many LD blocks were observed in Figure S3, inferring high LD between some SSR loci.

Genetic diversity of sesame has been demonstrated using universal markers [8–13]. However, few co-dominant polymorphic markers, ranging from 10 to 50, were used in these studies [27,32]. Few polymorphic markers might not distinguish the samples accurately. In the present study, 218 novel

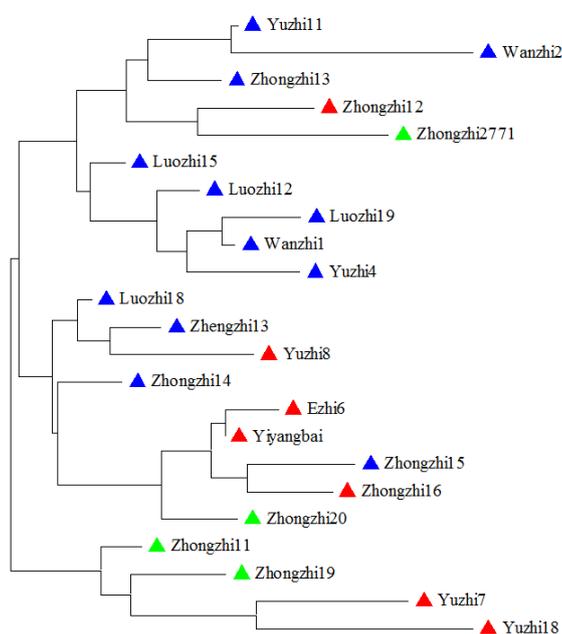
genomic SSRs were developed and used in the genetic diversity analysis. Thirty-one worldwide represent sesame accessions, including both landrace and elite, were collected and used to detect the genetic diversity and relationships. Population structure and phylogenetic analyses indicated that sesame landraces from China and other countries might be inconsistent and elites from China had several common ancient varieties. This was the first attempt at sesame genetic and genotypic diversity analysis using SSR markers developed from de novo sequencing.

2.4. Core SSR Marker Development

Among the 218 polymorphic SSR markers, 137 were be considered as high polymorphic SSR markers ($PIC > 0.6$), while only 102 were easy to identify, repeatable and highly polymorphic. Then, 32 SSR markers were randomly selected from the 102 SSR markers to construct phylogenetic tree of the 31 sesame samples. Finally, we obtained 32 SSR markers to build a UPGMA tree similar to that constructed by the 218 SSR markers (Figure S4). These 32 SSR markers were suggested to be core sesame SSR markers for genetic diversity detection. All of these markers are indicated in Table S2. The core SSR markers were evenly distributed in 16 linkage groups (LGs) of the sesame genome, including two markers in each LG.

Twenty additional sesame elites which widely planted in central China were selected. With the three sesame varieties in the 31 sesame samples, we collected 23 elites (Table S4). These elites could be divided into three groups since they had three different major parent varieties (Yiyangbai, Yuzhi4 and Zhongzhi11). To determine if the core SSR markers could reveal sesame varieties genealogy, core SSR markers were used to screen the 23 sesame varieties and construct a phylogenetic tree. Figure 5 shows the relationship among these accessions. Although the varieties were not strictly clustered into three groups, all progenitors tended to be in nearby clades. The results indicated that the 32 core SSR markers could be a useful tool for genetic diversity detection of sesame varieties.

Figure 5. Dendrogram of 23 sesame varieties based on 20 high polymorphism markers. Samples are listed in Table S4.



Many SSR markers are typically used for the genetic diversity detection. However, the process is time-consuming for a large number of samples. Core SSR markers have been developed and utilized in some crops, such as soybean [33], maize [34], cotton [35] and rapeseed [36]. Core SSR markers are a set of highly polymorphic SSR markers which selected from plenty genomic SSRs and maximally reflect the diversity and relationship among varieties or germplasm. However, sesame core SSR marker development has not been reported. In this study, 32 core SSR markers were selected from 218 polymorphic SSR markers and could highly reflect sesame variety genealogy.

3. Experimental

3.1. Plant Materials, Genome Sequencing, and Genome Survey

The sample used for whole-genome de novo sequencing was “Zhongzhi13”, an elite sesame cultivar which was introduced to most of the major sesame planting areas over the last 10 years. In total, eight paired-end sequencing libraries with insert sizes of approximately 180 bp, 500 bp, 800 bp, 2 kb, 5 kb, 10 kb and 20 kb were constructed and sequenced using the Illumina Hiseq 2000 platform.

Thirty-one accessions of sesame collected from worldwide (Table S3), were used to evaluate the SSR suitability for genetic distance analysis. Young leaves were collected and frozen in liquid nitrogen prior to genomic DNA extraction using CTAB methods [37]. DNA concentrations were measured spectrophotometrically at 260 nm, and the extracts were electrophoresed on 1% agarose to confirm the quality. The purified DNAs was standardized at 40 ng/μL and stored at −40 °C.

3.2. SSR Identification and Primer Design

The Perl script MICROSATELLITE (MISA) was used to identify microsatellites in sesame genomes [38]. To identify the presence of SSRs, only two to six nucleotide motifs were considered, and the minimum repeat unit was defined as six reiterations for dinucleotides, and five reiterations for other repeat units. Mononucleotide repeats and complex SSR types were excluded from the study since distinguishing genuine mononucleotide repeats from polyadenylation products and single nucleotide stretch errors generated by sequencing was difficult.

The SSR primers were designed using BatchPrimer3 interface modules [39]. We selected 1,500 primers that met the following parameters: 100–300 final product length (optimal 200 bp), primer size from 19 to 27 bp (optimal 23 bp) and GC content 40%–70% (optimal 50%); the annealing temperature was set at 50–60 °C (optimal 55 °C).

Primers were synthesized by Invitrogen Trading (Shanghai) Co., Ltd. We primarily tested four cultivars (Aizhima, Mishuozhima, Gaozhima and Silengzhima) for 1,500 SSR loci by PAGE to confirm their suitability.

3.3. PCR and Gel Electrophoresis

Each 20 μL reaction mixture contained 10× PCR buffer (plus Mg²⁺), 0.2 mM of each dNTP, 5 pmol of each reverse, 4 pmol of the tail primer, 1 pmol of the forward primer, 0.5 units of rTaq polymerase (TaKaRa, Dalian, China) and 40 ng genomic DNA template. Each primer pair had a 20-bp interval according to the expected amplicon size.

DNA amplification was accomplished in a PCR system (Mastercycler: Eppendorf, Hamburg, Germany) programmed at 94 °C for 5 min for initial denaturation, then 35 cycles at 94 °C (30 s)/55 °C (30 s)/72 °C (30 s), followed by the extension step for 10 min at 72 °C. Each PCR product was run on 1% agarose gel at 110 V for a quality check. Subsequently, the PCR products were run on 6% denaturing polyacrylamide gel at 80 W constant gain and polymorphism was detected by silver staining [40].

3.4. Data Analysis

The number of alleles (N_A), observed heterozygosity (H_o) and expected heterozygosity (H_e) were calculated using POPGENE version 1.32 [41]. Major allele frequency (M_{AF}), chi-square test for HWE allele frequencies (P_{HW}) and PIC were calculated using PowerMarker version 3.25 [42]. Genetic similarity among the 31 sesame accessions was calculated according to Dice's coefficients using Nei's coefficient index with the FreeTree 0.9.1.50 software [43], and phylogenetic trees were constructed using the NJ and UPGMA method with 1000 bootstrap test [44]. A Bayesian clustering analysis of the samples was constructed by Structure version 2.2 [45]. Five runs of STRUCTURE were carried out for each population number (K : from 1 to 10), and each run started with 10,000 burn-ins, followed by 100,000 iterations. LD level of SSR loci in sesame was tested by Tassel 3.0 [46]. Thirty-two high polymorphism SSR markers were selected as core SSR markers. Twenty-three Chinese sesame varieties were screened using these core SSR markers. The Chinese sesame variety dendrograms were constructed by genetic distance.

4. Conclusions

Based on the first sequenced and assembled sesame genome, genomic SSR markers were surveyed in the present study. Genome survey identified a total of 23,438 SSR loci from the sesame genome. However, sesame genome SSR frequency was less than that of *Arabidopsis*, rice and sorghum. Different motif frequencies decreased stepwise with increased motif length, and dinucleotide repeats dominated the SSR types. Among the dinucleotide repeat motifs, (AT) n was the most common repeat unit. One thousand five hundred SSR primer pairs were designed and synthesized, and 218 of them were identified as polymorphic markers. Polymorphic SSR markers were used to investigate genetic diversity among 31 worldwide sesame accessions. The population structure and NJ analyses separated all samples into two clusters that corresponded to the geographical areas from which they originated. Thirty-two SSR markers were selected as the core SSR markers and successfully used in genealogy analysis of 23 elites. These SSR markers could provide a useful tool for genetic linkage map construction, genetic diversity detection and marker-assisted selection breeding of sesame.

Supplementary Materials

Supplementary materials can be accessed at: <http://www.mdpi.com/1420-3049/19/4/5150/s1>.

Acknowledgments

This work was supported by the National Basic Research Program of China (973 Program) (No. 2011CB109304), the National Natural Science Foundation of China (No. 31271766, No. 31101182 and No. 31201242), and China Agriculture Research System (No. CARS-15).

Author Contributions

Z.X.R., W.X., W.L.H. designed the experiments. Z.X.R., Z.Y.X., Q.X.Q. collected plant materials. W.X., W.L.H., W.X.L., D.X. and Z.J. performed the SSR experiments and analyzed the data. W.X. drafted this manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Nayar, N.M.; Mehra, K.L. Sesame: Its uses, botany, cytogenetics, and origin. *Econ. Bot.* **1970**, *24*, 20–31.
2. Johnson, L.A.; Suleiman, T.M.; Lusas, E.W. Sesame protein: A review and prospectus. *J. Am. Oil Chem. Soc.* **1979**, *56*, 463–468.
3. Wei, W.; Zhang, Y.; Lv, H.; Li, D.; Wang, L.; Zhang, X. Association analysis for quality traits in a diverse panel of Chinese sesame (*Sesamum indicum* L.) germplasm. *J. Integr. Plant Biol.* **2013**, *55*, 745–758.
4. Yen, G.C. Influence of seed roasting process on the changes in composition and quality of sesame (*Sesame indicum*) oil. *J. Sci. Food Agric.* **1990**, *50*, 563–570.
5. Cheung, S.C.; Szeto, Y.T.; Benzie, I.F. Antioxidant protection of edible oils. *Plant Foods Hum. Nutr.* **2007**, *62*, 39–42.
6. Hirata, F.; Fujita, K.; Ishikura, Y.; Hosoda, K.; Ishikawa, T.; Nakamura, H. Hypocholesterolemic effect of sesame lignan in humans. *Atherosclerosis* **1996**, *122*, 135–136.
7. Tsai, C.M.; Chen, P.R.; Chien, K.L.; Su, T.C.; Chang, C.J.; Liu, T.-L.; Cheng, H. Dietary sesame reduces serum cholesterol and enhances antioxidant capacity in hypercholesterolemia. *Nutr. Res.* **2005**, *25*, 559–567.
8. Bhat, K.V.; Babrekar, P.P.; Lakhanpaul, S. Study of genetic diversity in Indian and exotic sesame (*Sesamum indicum* L.) germplasm using random amplified polymorphic DNA (RAPD) markers. *Euphytica* **1999**, *110*, 21–33.
9. Ercan, A.G.; Taskin, M.; Turgut, K. Analysis of genetic diversity in Turkish sesame (*Sesamum indicum* L.) populations using RAPD markers. *Genet. Res. Crop Evol.* **2004**, *51*, 599–607.
10. Kim, D.H.; Zur, G.; Danin-Poleg, Y.; Lee, S.W.; Shim, K.B.; Kang, C.W.; Kashi, Y. Genetic relationships of sesame germplasm collection as revealed by inter-simple sequence repeats. *Plant Breed.* **2002**, *121*, 259–262.

11. Hernan, E.L.; Petr, K. Genetic relationship and diversity in a sesame (*Sesamum indicum* L.) germplasm collection using amplified fragment length polymorphism (AFLP). *BMC Genet.* **2006**, *7*, 10.
12. Zhang, Y.X.; Zhang, X.R.; Hua, W.; Wang, L.H.; Che, Z. Analysis of genetic diversity among indigenous landraces from sesame (*Sesamum indicum* L.) core collection in China as revealed by SRAP and SSR markers. *Genes Genomics* **2010**, *32*, 207–215.
13. Wei, L.B.; Zhang, H.Y.; Zheng, Y.Z.; Guo, W.Z.; Zhang, T.Z. Developing EST-derived microsatellites in sesame (*Sesamum indicum* L.). *Acta Agron. Sin.* **2008**, *34*, 2077–2084.
14. Watanabe, K.N.; Hirano, R.; Ishii, H.; Oo, T.H.; Gilani, S.A.; Kikuchi, A. Propagation management methods have altered the genetic variability of two traditional mango varieties in Myanmar, as revealed by SSR. *Plant Genet. Resour. C* **2011**, *9*, 404–410.
15. Chen, C.; Zhou, P.; Choi, Y.; Huang, S.; Gmitter, F. Mining and characterizing microsatellites from citrus ESTs. *Theor. Appl. Genet.* **2006**, *112*, 1248–1257.
16. Fraser, L.G.; Harvey, C.F.; Crowhurst, R.N.; de Silva, H.N. EST-derived microsatellites from *Actinidia* species and their potential for mapping. *Theor. Appl. Genet.* **2004**, *108*, 1010–1016.
17. Jiao, Y.; Jia, H.; Li, X.; Chai, M.; Jia, H. Development of simple sequence repeat (SSR) markers from a genome survey of Chinese bayberry (*Myrica rubra*). *BMC Genomics* **2012**, *13*, 201.
18. Kresovich, S.; Szewc-McFadden, A.K.; Bliet, S.M.; McFerson, J.R. Abundance and characterization of simple-sequence repeats (SSRs) isolated from a size-fractionated genomic library of *Brassica napus* L. (rapeseed). *Theor. Appl. Genet.* **1995**, *91*, 206–211.
19. Hopkins, M.S.; Casa, A.M.; Wang, T.; Mitchell, S.E.; Dean, R.E.; Kochert, G.D.; Kresovich, S. Discovery and characterization of polymorphic simple sequence repeats (SSRs) in peanut. *Crop Sci.* **1999**, *39*, 1243–1247.
20. Akkaya, M.S.; Bhagwat, A.A.; Cregan, P.B. Length polymorphisms of simple sequences repeat DNA in soybean. *Genetics* **1992**, *132*, 1131–1139.
21. Wei, W.; Qi, X.; Wang, L.; Zhang, Y.; Hua, W.; Li, D.; Lv, H.; Zhang, X. Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. *BMC Genomics* **2011**, *12*, 451.
22. Zhang, H.; Wei, L.; Miao, H.; Zhang, T.; Wang, C. Development and validation of genic-SSR markers in sesame by RNA-seq. *BMC Genomics* **2012**, *13*, 316.
23. Wu, Y.Q.; Wang, Y.W.; Samuels, T.D. Development of 1030 genomic SSR markers in switchgrass. *Theor. Appl. Genet.* **2011**, *122*, 677–686.
24. Wang, L.; Yu, S.; Tong, C.; Zhao, Y.; Liu, Y.; Song, C.; Zhang, Y.; Zhang, X.; Wang, Y.; Hua, W.; *et al.* Genome sequencing of the high crop sesame provides insight into oil biosynthesis. *Genome Biol.* **2014**, *15*, R39.
25. Sonah, H.; Deshmukh, R.K.; Sharma, A.; Singh, V.P.; Gupta, D.K.; Gacche, R.N.; Rana, J.C.; Singh, N.K.; Sharma, T.R. Genome-wide distribution and organization of microsatellites in plants: An insight into marker development in *Brachypodium*. *PLoS One* **2011**, *6*, e21298.
26. Cheng, X.; Xu, J.; Xia, S.; Gu, J.; Yang, Y.; Fu, J.; Qian, X.; Zhang, S.; Wu, J.; Liu, K. Development and genetic mapping of microsatellite markers from genome survey sequences in *Brassica napus*. *Theor. Appl. Genet.* **2009**, *118*, 1121–1131.

27. Wang, L.; Zhang, Y.; Qi, X.; Gao, Y.; Zhang, X. Development and characterization of 59 polymorphic cDNA-SSR markers for the edible oil crop *Sesamum indicum* (Pedaliaceae). *Am. J. Bot.* **2012**, *99*, e394–e398.
28. McCouch, S.R.; Cho, Y.G.; Ishii, T.; Temnykh, S.; Chen, X.; Lipovich, L.; Park, W.D.; Ayres, N.; Cartinhour, S. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* **2000**, *100*, 713–722.
29. Yermanos, D.M. Sesame. In *Hybridization of Crop Plants*; Fehr, W.R., Hadley, H.H., Eds.; American Society of Agronomy: Crop Science Society of America: Madison, MI, USA, 1980; pp. 25–36.
30. Flint-Garcia, S.A.; Thornsberry, J.M.; Buckler, I.V. Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* **2003**, *54*, 357–374.
31. Remington, D.L.; Thornsberry, J.M.; Matsuoka, Y.; Wilson, L.M.; Whitt, S.R.; Doebley, J.; Kresovich, S.; Goodman, M.M.; Buckler, E.S. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 11479–11484.
32. Dixit, A.; Jin, M.H.; Chung, J.W.; Yu, J.W.; Chung, H.K.; Ma, K.-H.; Park, Y.-J.; Cho, E.-G. Development of polymorphic microsatellite markers in sesame (*Sesamum indicum* L.). *Mol. Ecol. Notes* **2005**, *5*, 736–738.
33. Hua, X.; Chang, R.; Cao, Y.; Zhang, M.; Feng, Z.; Qiu, L. Selection of core SSR loci by using Chinese autumn soybean. *Sci. Agric. Sin.* **2006**, *36*, 360–366.
34. Wang, F.; Yi, H.; Zhao, J.; Wang, L.; Sun, Y. Construction of a common PCR program for SSR core primer in maize. *J. Maize Sci.* **2008**, *16*, 19–21.
35. Pan, Z.; Sun, J.; Wang, X.; Jia, Y.; Zhou, Z.; Pang, B.; Du, X.M. Screening of SSR core primers with polymorphism on a cotton panel. *Biodivers. Sci.* **2008**, *16*, 555–561.
36. Li, H.; Jun, Y.; LV, Z.; Yi, B.; Wen, J.; Fu, T.-D.; Tu, J.-X.; Ma, C.-Z.; Shen, J.-X. Screening of *Brassica napus* core SSR primers. *Chin. J. Oil Crop Sci.* **2010**, *32*, 329–336.
37. Wei, X.; Wang, R.S.; Cao, L.R.; Yuan, N.N.; Huang, J.; Qiao, W.-H.; Zhang, W.-X.; Zeng, H.-L.; Yang, Q.W. Origin of *Oryza sativa* in China inferred by nucleotide polymorphisms of organelle DNA. *PLoS One* **2012**, *7*, e49546.
38. Thiel, T.; Michalek, W.; Varshney, R.K.; Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **2003**, *106*, 411–422.
39. Kortt, A.A.; Caldwell, J.B.; Lilley, G.G.; Higgins, T.J.V. Amino acid and cDNA sequences of a methionine-rich 2S protein from sunflower seed (*Helianthus annuus* L.). *Eur. J. Biochem.* **1991**, *195*, 329–334.
40. Bassam, B.J.; Caetano-Anolles, G.; Gresshoff, P.M. Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal. Biochem.* **1991**, *196*, 80–83.
41. Yeh, F.C.; Boyle, T.J.B. Population genetic analysis of co-dominant and dominant markers and quantitative traits. *Belg. J. Bot.* **1997**, *129*, 157.
42. Liu, K.J.; Muse, S.V. PowerMarker, an integrated analysis environment for genetic marker analysis. *Bioinformatics* **2005**, *21*, 2128–2129.

43. Hampl, V.; Pavlicek, A.; Flegr, J. Construction and bootstrap analysis of DNA fingerprinting-based phylogenetic trees with the freeware program FreeTree, application to trichomonad parasites. *Int. J. Syst. Evol. Microbiol.* **2001**, *51*, 731–735.
44. Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5, Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **2011**, *28*, 2731–2739.
45. Falush, D.; Stephens, M.; Pritchard, J.K. Inference of population structure, extensions to linked loci and correlated allele frequencies. *Genetics* **2003**, *164*, 1567–1587.
46. Bradbury, P.J.; Zhang, Z.; Koon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635.

Sample Availability: Samples are available from the authors' institute.

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).