Molecular and Cellular Probes 30 (2016) 153–160

Contents lists available at ScienceDirect

Molecular and Cellular Probes

journal homepage: www.elsevier.com/locate/ymcpr

Original research article

Genome-wide identification and validation of simple sequence repeats (SSRs) from *Asparagus officinalis*

Shufen Li [a], Guojun Zhang [b], Xu Li [a], Lianjun Wang [c], Jinhong Yuan [a], Chuanliang Deng [a], Wujun Gao [a, *]

[a] College of Life Sciences, Henan Normal University, Xinxiang 453007, Henan, PR China
[b] School of Basic Medical Sciences, Xinxiang Medical University, Xinxiang 453003, Henan, PR China
[c] Institute of Food Corps, Hubei Academy of Agricultural Sciences, Wuhan 430064, Hubei, PR China

ARTICLE INFO

Article history:
Received 20 January 2016
Received in revised form
26 February 2016
Accepted 8 March 2016
Available online 14 March 2016

Keywords:
Asparagus officinalis
Genetic diversity
Genomic SSRs
Microsatellites
Motif

ABSTRACT

Garden asparagus (*Asparagus officinalis*), an important vegetable cultivated worldwide, can also serve as a model dioecious plant species in the study of sex determination and sex chromosome evolution. However, limited DNA marker resources have been developed and used for this species. To expand these resources, we examined the DNA sequences for simple sequence repeats (SSRs) in 163,406 scaffolds representing approximately 400 Mbp of the *A. officinalis* genome. A total of 87,576 SSRs were identified in 59,565 scaffolds. The most abundant SSR repeats were trinucleotide and tetranucleotide, accounting for 29.2 and 29.1% of the total SSRs, respectively, followed by di-, penta-, hexa-, hepta-, and octanucleotides. The AG motif was most common among dinucleotides and was also the most frequent motif in the entire *A. officinalis* genome, representing 14.7% of all SSRs. A total of 41,917 SSR primers pairs were designed to amplify SSRs. Twenty-two genomic SSR markers were tested in 39 asparagus accessions belonging to ten cultivars and one accession of *Asparagus setaceus* for determination of genetic diversity. The intra-species polymorphism information content (PIC) values of the 22 genomic SSR markers were intermediate, with an average of 0.41. The genetic diversity between the ten *A. officinalis* cultivars was low, and the UPGMA dendrogram was largely unrelated to cultivars. It is here suggested that the sex of individuals is an important factor influencing the clustering results. The information reported here provides new information about the organization of the microsatellites in *A. officinalis* genome and lays a foundation for further genetic studies and breeding applications of *A. officinalis* and related species.

1. Introduction

Microsatellites, or simple sequence repeats (SSRs), are short tandemly repeated DNA sequences of 1–6 nucleotides that occur ubiquitously in the genomes of prokaryotic and eukaryotic organisms [1,2]. The number of repeat units may vary among individual genotypes due to their high mutation rate, making SSRs a rich source of highly polymorphic markers [3]. Among the various molecular markers, such as random amplified polymorphic DNA (RAPD) [4], inter-simple sequence repeat (ISSR) [5], amplified fragment length polymorphism (AFLP) [6], SSR markers are especially useful because they are mostly co-dominant, abundant in genomes, randomly distributed within the genome, producible, and highly polymorphic [7–9].

Currently, SSR markers have been widely used for genetic linkage mapping [10], genetic diversity assessment [11,12], marker-assisted selection [13,14], population genetics [15–17], and evolutionary studies [18]. In addition to their usefulness as genetic markers, a growing number of studies have shown that microsatellites play important roles in many biological processes, such as organization of the genome and chromatin [19], regulation of DNA metabolism [20,21], regulation of gene expression [22,23], and genome evolution [24,25].

Garden asparagus (*Asparagus officinalis*), which belongs to the genus *Asparagus*, is a dioecious species (2n = 2x = 20) with a haploid genome size of 1308 Mb [26,27]. Because it is an economically and nutritionally important vegetable crop cultivated worldwide, the breeding of *A. officinalis* with favorable traits such as high yield, disease resistance, and stress resistance has great commercial value. The development of molecular markers in

* Corresponding author.
E-mail address: gaowujun1@163.com (W. Gao).

http://dx.doi.org/10.1016/j.mcp.2016.03.003
0890-8508/

A. officinalis has potential uses in marker-assisted selection (MAS) during breeding. Abundant genome-wide molecular markers are necessary for MAS. Currently, the number of molecular markers, particularly SSR markers for *A. officinalis* is still limited [28,29]. Large-scale development of SSR markers is needed for genetic mapping and marker-assisted selection studies. *A. officinalis* is also used as a model dioecious plant species for the study of sex determination and the evolution of sex chromosomes [30–32]. Reports have shown that microsatellites may play important roles in the evolution of sex chromosomes [24,33,34]. In this way, investigation of the distribution and variation in microsatellites in the genome of *A. officinalis* could lay a foundation for further study of the evolution of sex chromosomes with respect of microsatellites.

Identification of genomic SSRs and subsequent conversion to markers was once expensive and time-consuming [35]. Currently, the process of genomic SSR mining and marker development is speedy and low cost. This is due to the advent of next-generation sequencing technologies. This work concerning genome-wide analysis of SSRs and validation based on nest-generation sequencing data have been performed successfully in many plant species [7,36–38]. Genome-wide analysis of SSRs, coupled with experimental validation may provide insight into genome organization and provide abundant markers for genetic, genomic, and evolutionary studies.

Recently, we performed Illumina HiSeq2000 sequencing of the *A. officinalis* genome. A total of 17 Gb of sequence (12 × coverage) were generated and assembled into 163,406 scaffolds with a total cumulative length of about 400 Mbp, which represents about 30% of the *A. officinalis* genome [39]. The availability of sequencing data provides opportunities to generate large numbers of SSR markers in *A. officinalis* faster and at a lower cost. Thus, in this study, using the Illumina sequencing data, we conduct a large-scale genome-wide characterization of SSR loci was conducted; and the usefulness of SSR loci for study of the genetic diversity of *A. officinalis* cultivars was assessed. This research may provide a helpful reference and effective SSR marker resources for comprehensive study of *A. officinalis* germplasms, and may also aid in future breeding of *A. officinalis*.

2. Materials and methods

2.1. Plant materials and DNA extraction

Ten *A. officinalis* cultivars (Table 1) were used for validation of SSR markers and examination of genetic diversity. *Asparagus setaceus*, a close relative of *A. officinalis*, here served as an outgroup for dendrogram analysis. All plants were grown in the garden field of Henan Normal University. Total genomic DNA was extracted from the leaves of *A. officinalis* and *A. setaceus* using a traditional CTAB method [40].

Table 1
Plant materials used in this study.

Number	Cultivar/species name	Number of individuals
1	TD818	4 (2 female and 2 male)
2	UC309	4 (2 female and 2 male)
3	MLJZ	4 (2 female and 2 male)
4	Jersey Giant	4 (2 female and 2 male)
5	Apollo	4 (2 female and 2 male)
6	Grande	4 (2 female and 2 male)
7	U88	4 (2 female and 2 male)
8	UC800	4 (2 female and 2 male)
9	UC157	4 (2 female and 2 male)
10	UC308	3 (1 female and 2 male)
11	*A. setaceus*	1

2.2. Detection of genomic SSRs and primer design

A large-scale, genome-wide SSR search was performed using assembled scaffolds based on Illumina sequencing. Perl scripting source code here served as a stand-alone SSR identification tool (SSRIT [41]). It was downloaded from http://www.gramene.org/db/markers/ssrtool and modified to run in a batch mode. Both perfect and compound repeats of basic motifs ranging from 2- to 8-bp of SSRs were considered. The minimum repeat unit was defined as six reiterations for dinucleotides, four reiterations for trinucleotides, and three reiterations for other repeat units. Oligonucleotide primers were designed using the SSR flanking sequences using Primer 3 software [42]. The primers met the following parameters: 100–300 amplicons (optimal 200 bp); primer length from 18 to 25 bp (optimal 22 bp), GC content 40–60% (optimum 50%); Tm value 55–62 °C (optimal 59 °C).

2.3. SSR amplification

SSR amplification was performed in 25 μL PCR reaction mixture containing 2.5 μL 10 × PCR buffer, 50 ng genomic DNA, 0.5 U Taq polymerase (Takara, Dalian, China), 0.2 μM of each primer, and 200 μM dNTPs (Takara). The reaction was denatured at 94 °C for 2 min and then subjected to 35 cycles of 94 °C for 1 min, annealing temperature (based on each pair of the primers) for 30 s, 72 °C for 1 min and finally incubated at 72 °C for 10 min. The PCR products were resolved on 8% nondenatured polyacrylamide gels and visualized using silver staining. The band size is reported using a DL500 marker (Takara) as the reference point. The polymorphism was determined according to the presence or absence of the SSR band.

2.4. Data analysis

Individual bands amplified by SSR primers in SSR banding profile were scored as presence (1) or absence (0). Low-frequency fragments (<4 occurrences) were removed from the data set because these fragments might be unreliable. The statistical analysis of genetic variations was performed using POPGENE32 (Version 1.31) software [43]. UPGMA (Unweighted Pair Group Method with Arithmetic Mean) dendrogram was constructed using MEGA 6 [44] and Nei's unbiased genetic distance. The polymorphic information content (PIC) values for each marker were calculated according to the following formula: $PIC = 1 - \sum_{i=1}^{n} P_{ij}^2$, where p_{ij} is the frequency of the *j*th pattern for SSR marker i.

3. Results

3.1. Detection of asparagus genomic SSR loci

Perfect SSRs with ≥3 repeat units and a minimum total length of 12 bp were detected in ≈400 Mbp of non-redundant *A. officinalis* genomic sequence, representing about 30% of its 1308 Mbp nuclear genome. Analysis showed that 59,565 (36%) sequences out of the 163,406 scaffolds [39] contained SSRs; 17,525 (29%) of the sequences contained more than one SSR locus. A total of 87,576 SSRs with perfect repeats were detected in the *A. officinalis* genome (Fig. 1). On average, the overall SSR density across the genome was 218.9 SSRs/Mbp, i.e., one SSR every 4.6 kb of sequence, excluding mononucleotide SSRs. Taken together, the total length of di-to octanucleotide SSR sequences was about 1350 kb, accounting for about 0.34% of the assembled *A. officinalis* genome.