Full Length Research Paper

Isolation and characterization of polymorphic microsatellite markers from flax (*Linum usitatissimum* L.)

Xin Deng¹, SongHua Long¹, DongFeng He², Xiang Li², YuFu Wang¹, DongMei Hao¹, CaiSheng Qiu¹ and XinBo Chen^{1, 2}*

¹Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, No.348 West Xianjia Lake Road, Changsha, Hunan Province, 410205, China.

²Hunan Provincial Key Laboratory of Crop Germplasm Innovation and Utilization, Hunan Agricultural University, Changsha, Hunan Province, 410128, China.

Accepted 21 December, 2010

Flax (*Linum usitatissimum* L.) is the third largest natural fiber crop and one of the five major oil crops in the world. Flax can be used as model plant for the bast fiber plants because of its small size and self pollination characters. Development of highly polymorphic co-dominant markers will be important for the molecular and genomic research in flax. Two microsatellite-enriched flax genomic libraries were constructed for trinucleotide TTC and ATC motifs. A total of 206 new microsatellite-containing sequences were identified and classified. Thirty eight polymorphic microsatellite markers were characterized and evaluated in eight cultivars from different countries and regions. These loci produced 2 to 12 alleles per locus with an average of 3.395. The observed and expected heterozygosities ranged from 0.000 to 1.000 (mean 0.257) and from 0.125 to 0.950 (mean 0.516), respectively. The mean polymorphic information content (PIC) value over 38 loci was 0.429, with 13 loci having PIC greater than 0.5. These novel polymorphic microsatellite loci will be useful in genetic linkage map construction, germplasm classification and identification, gene identification and QTL mapping, and marker-assisted selection in breeding of *L. usitatissimum*.

Key words: Flax, *Linum usitatissimum*, microsatellite, polymorphism.

INTRODUCTION

Flax (*Linum usitatissimum* L.), also called common flax or linseed, is an annual herb, which is the third largest natural fiber crop and one of the five major oil crops in the world. Flax is a small size and self pollination herb that has been thought to be the model plant for the bast fiber plants. At present, fiber flax cultivars are mainly grown

in some regions of northern Europe, Russia and China, while distinct linseed flax varieties are widely grown in cool temperate regions of Argentina, India, China, Russia, the USA and Canada (Millam et al., 2005).

Different molecular marker techniques have been applied in the flax molecular marker development and in flax genetic resource evaluation. These include random amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP) and simple sequence repeat (SSR) (Spielmeyer et al., 1998; Oh et al., 2000; Fu et al., 2002, 2003a, b; McBreen et al., 2003; Adugna et al., 2006; Roose-Amsaleg et al., 2006). However, the numbers of effective flax molecular markers were still limited, and some marker types (RAPD, RFLP and AFLP) are not easily reproducible or quite laborious. Microsatellites, also known as simple sequence repeats (SSRs) are considered

^{*}Corresponding author. E-mail: xinbochen@live.cn. Fax: +86-731-88998528.

Abbreviations: SSR, Simple sequence repeat; RAPD, random amplified polymorphic DNA; RFLP, restriction fragment length polymorphism; AFLP, amplified fragment length polymorphism; ESTs, expressed sequence tags; H_0 , observed heterozygosities; H_E , expected heterozygosities; PIC, polymorphic information content.

Code	Cultivar	Origin	Plant Type
1	Baihua	France	Fibre
2	Shanxidatong	Shanxi, china	Linseed
3	NDR714	America	Linseed
4	WINONA SEL	America	Linseed
5	K-1195	Ethiopia	Fibre
6	CDC BETHUE	Canada	Linseed
7	SOMME	Canada	Linseed
8	AC WATSON	Canada	Linseed

Table 1. Description of the flax cultivars used for polymorphism analysis.

as ideal markers for their being highly polymorphic, codominant, and comparatively simple and inexpensive (Powell et al., 1996). Large numbers of microsatellite markers had been developed in other plants, such as soybean, cotton and rice. However, microsatellite marker development in flax lags behind those in other important cash crops. Microsatellite markers can be derived from either genomic sequences or expressed sequence tags (ESTs). In earlier study, 248 EST-SSR loci were developed from flax by Sylvie Cloutier et al., (2009). For the genomic sequences, only Roose-Amsaleg et al. (2006) and Deng et al. (2010) reported the development of 28 and 35 SSR markers, and only two trinucleotide motifs were included. In this study, we constructed two microsatellite-enriched flax genomic libraries and isolated 206 new microsatellite-containing sequences of trinucleotide repeats, and 38 polymorphic microsatellite markers were developed.

MATERIALS AND METHODS

Plant materials and DNA extraction

Eight flax cultivars from different countries and regions (Table 1) were grown in a growth cabinet maintained at 22/19°C and 70% humidity. DNA was extracted from the seedling leaves using the TIAN Gel Midi Purification Kit (Tiangen).

Microsatellite-enriched library construction

Two microsatellite enriched-libraries of flax were built using biotinlabelled microsatellite oligoprobes [(TTC)10 and (ATC)10] and streptavidin-coated magnetic beads (Carol et al., 2003; Lunt et al., 1999; Deng et al., 2008). Flax genomic DNA of cultivar Baihua was digested with restriction enzyme Mse I (Fermentas) and ligated to double-stranded linkers {primer A (5'-AGATGGAATTCGTACAC TCGT-3') and primer B (5'-TAACGAGTGTACGAATTCCATCT-3')}. The enrichment of the microsatellite containing DNA was performed according to Deng et al (2008). Briefly, the Linker ligated DNA was denatured and hybridized to biotinylated oligonucleotide probes (TTC)₁₀ or (ATC)₁₀ mixes, which were then captured on magnetic streptavidin beads (Dynal). Unhybridized DNA was washed away and the remaining microsatellite-enriched DNA was eluted from the beads. The enriched single-stranded microsatellite containing DNA fragments were amplified by polymerase chain reactions (PCR) using the primer A as a primer. The amplified products were ligated into pMD18-T vector (TaKaRa). Then the ligation mixture was transformed into *Escherichia coli* Top10 competent cells to form the microsatellite sequence enriched libraries.

Library screening and sequence analysis of the positive SSR containing clones

The PCR screening of positive SSR containing clones was performed essentially according to Deng et al. (2008), except that the primer pairs were primer A and microsatellite specific primer E (VRV (TTC)₁₀) or primer F (VRV (ATC)₁₀). While any recombinant with the adapters at both ends of the inserted fragments will have a band of the insert size using the primer A as primer, the presence of extra smaller bands may indicate existence of microsatellite sequence in the clone. These bands should be the product of primer A and primer E/ primer F. The positive SSR containing clones were selected and sequenced by Beijing Genomics Institute Co., Ltd. All sequences were analyzed using Bioedit Sequence Alignment Editor software (http://www.mbio.ncsu.edu/BioEdit/BioEdit.html). The repeat numbers were determined using software SSRHunter1.3 (Li and Wan, 2005). PCR primer pairs were designed using Primer 3 software (Rozen and Skaletsky, 2000).

PCR amplification of microsatellite markers

The designed microsatellite primer pairs were tested for amplification on eight cultivars from different countries and regions (2 of fiber flax and 6 of linseed, Table 1). PCR amplification was performed according to Deng et al. (2010). The products were separated on 8% polyacrylamide gels and visualized using silver nitrate stain. The results were documented with a Bio-Rad GelDoc2000 system. Molecular sizes of the polymorphic fragments were estimated against a 50 bp ladder (Tiangen).

Data analysis

For those polymorphic SSR primer pairs, the number of alleles, observed heterozygosities ($H_{\rm O}$), expected heterozygosities ($H_{\rm E}$) and polymorphic information content (PIC) were calculated using Cervus version 3.0.3 (Kalinowski et al., 2007). The PIC for each primer was calculated according to Cordeiro et al. (2003).

RESULTS AND DISCUSSION

Identification of microsatellites from the flax SSR enriched genomic libraries

Two flax libraries enriched for TTC and ATC micro-

736



Figure 1. Distribution of microsatellite classes (A) and number of repeat units (B) of the isolated non-redundant SSR containing sequences

satellite core motifs were developed. The TTC library yielded 1283 recombinant clones, while the ATC library yielded 696 clones. PCR screening of SSR containing clones identified 256 putative recombinants from the TTC using primer A and primer E (VRV (TTC)₁₀) and 191 putative recombinants from the ATC library using primer A and primer F (VRV $(ATC)_{10}$). Sequence analysis revealed that microsatellite motifs were present in 191 of the TTC recombinants and in 72 of the ATC recombinants, accounting for 14.9 and 10.3% of the total clones, respectively. Sequence alignment analysis was performed to remove redundant sequences. The average level of redundancy was 21.67%, with a higher redundancy in the TTC library (27.23%) and a lower redundancy in the ATC library (6.94%). A total of 206 non-redundant microsatellite-containing sequences were GenBank (accession numbers: submitted to the

GQ461360-GQ461565). The enrichment procedure was proved successful and efficient. The efficiency rates of other reported flax microsatellite-enriched genomic libraries were 7.26 for CT and GT (Deng et al., 2010), and 5.91% for TG and AAG (Roose-Amsaleg et al., 2006). Except for the efficiency of enrichment procedure, the difference in enrichment efficiency probably also reflect difference in microsatellite repeat type representation. Cloutier et al. (2009) found in the flax EST-SSR that trinucleotide motif TTC was also the most abundant motif.

On the basis of the organization of the identified repeat motifs, the microsatellite sequences were classified as 113 perfect, 26 compound and 67 interrupted. The average numbers of repeat units were 18 and 10 for TTC and ATC, respectively (Figure 1).

We could design successful primer pairs from 92 of the

Locus	Repeat motif	Primer sequences (5'- 3')	T _a (℃)	k	Allele size (bp)	Ho	H _E	PIC Value	Accession No
LU1	(TTC) ₃₄	F: TCATTCATCTCCTTCCACTAAAA R: TTGAAAGCCCTAGTAGACACCA	58	4	146-179	0.000	0.733	0.630	GQ461360
LU2	(TTC) ₁₁	F: TCCGGACCCTTTCAATATCA R: AACTACCGCCGGTGATGA	60	3	139-148	0.000	0.567	0.468	GQ461362
LU3	(TTC) ₁₁	F: GCTCGTGATCTCCTTCATCC R: AAAACCACGTCCAGATGCTC	60	4	153-162	0.125	0.642	0.525	GQ461364
LU4	(TTC) ₁₀	F: TTATTTCCGGACCCTTTCAA R: AAACTACCGCCGGTGATGAT	61	2	106-148	0.000	0.233	0.195	GQ461365
LU5	(TTC)7	F: GTCACTGGGTGTGTGTTTGC R: AGCAGAAGAAGATGGCGAAA	60	3	134-140	0.000	0.433	0.371	GQ461369
LU6	(TTC)7	F: CCCCATTTCTACCATCTCCTT R: CAACAGCGGAACTGATGAAA	60	3	125-137	0.000	0.567	0.468	GQ461373
LU7	(TTC) ₂₁	F: CATCCAACAAAGGGTGGTG R: GGAACAAAGGGTAGCCATGA	60	5	134-146	1.000	0.700	0.595	GQ461381
LU8	(TTC) ₁₂ TTT(TT C) ₂₂ TTT(TTC) ₇	F: TCCCGTAATATTCTATGTTCTTCC R: TGAGTTGGACCTTACAAGACTCA	58	12	144-228	1.000	0.950	0.881	GQ461382
LU9	(TTC) ₁₇	F: TTGCGTGATTATCTGCTTCG R: ATGGCAGGTTCTGCTGTTTC	60	4	102-150	0.000	0.700	0.605	GQ461384
LU10	(TTC) ₁₀	F: GCCTAAAGCTGATGCGTTTC R: TGTCAGGCTCCTTCTTTTGC	60	5	141-159	0.750	0.792	0.701	GQ461385
LU11	(TTC) ₂₁	F: ATGGCAGGTTCTGCTGTTTC R: TTGCGTGATTATCTGCTTCG	60	5	105-153	0.250	0.708	0.618	GQ461393
LU12	(TTC) ₄	F: GGGATTGAGAAGAGGGCATA R: GTTGGGGTGAAGAGGAACAA	60	2	127-133	0.250	0.400	0.305	GQ461396
LU13	(TTC) ₁₂	F: AAGATGACGTCGGTGGTGAT R: CGGAACCTTCCATTTTCCTC	60	2	105-111	0.250	0.400	0.305	GQ461400
LU14	(TTC) ₆	F: GCTTGCGAGAAGAAGGAGAA R: TCACCAAAGGCATTCACAAA	60	2	141-147	0.125	0.325	0.258	GQ461401
LU15	(TTC) ₁₄	F: TGGACGACGATGAAGATGAA R: CCGCCGGGTACACTACTACT	60	3	108-114	0.750	0.633	0.511	GQ461409
LU16	(TTC) ₁₀	F: TTATTCTTGCCTGCCAATCG R: TCCAGCTCTTGCTCGTTCTT	61	2	145-148	0.125	0.125	0.110	GQ461410
LU17	(TTC) ₁₃	F: GCTGGACCTTACAAGCCTCA R: TTGGTGGGAGAACAACAAGA	60	3	144-150	0.125	0.508	0.427	GQ461413
LU18	(TTC)9	F: AGAGGCGGAGGGCATTAC R: TTGGAGAGTTGGAATCGAGA	59	4	139-145	0.375	0.517	0.443	GQ461418

 Table 2. Characteristics of the 38 microsatellite markers of L. usitatissimum.

206 sequences containing microsatellite repeats. The remaining sequences showed either high similarity or were not suitable for primer design because of short or missing flanking regions. Under optimized amplification conditions, two primer pairs had non-specific amplification products and 52 primer pairs did not show polymorphic amplification products. The remaining 38 primer pairs could detect at least one polymorphism among eight flax genotypes. The repeat motif, primer sequence and annealing temperature (Ta) of the 38 SSR markers are given in Table 2.

Characterization of microsatellite markers

The number of alleles, allele size variations, PIC, observed ($H_{\rm O}$) and expected ($H_{\rm E}$) heterozygosities of the 38 SSR markers among eight flax genotypes are showed in Table 2. Overall, 129 alleles were amplified, with an average of 3.395 alleles per locus and the number of allele ranging from 2 to 12 over 38 loci. Differences in the number of alleles were found in different microsatellite types. The average number of alleles was 3.679 and 2.600 for TTC and ATC markers, respectively. Our

Table 2. Continued.

LU19	(TTC) ₁₄		60	3	132-147	0.125	0.242	0.215	GQ461419
1.1100			00	0	100,100	0.000	0.500	0.050	00401401
L020	(TTC) ₁₅	R: CAAGAAGAGGCCCAGAATTG	60	2	108-123	0.000	0.500	0.359	GQ461421
11121	(TTC) (T	E: AAGGGTGGTGGTGGGGAAC	60	4	97-147	1 000	0 733	0.630	GO461422
2021	(TTC) ₁₈		00		07 117	1.000	0.700	0.000	GGIOTILL
11122		E: GATGGGGTTGAAGCCAGTAG	60	2	138-144	0.250	0 533	0 375	GO461424
LUZZ	(110)6		00	2	130-144	0.230	0.555	0.375	00401424
11100			60	0	100 100	0.000	0.000	0 105	00461405
L023	(110)7		00	2	133-130	0.000	0.233	0.195	GQ401425
11104			00	0	00.100	0.075	0.005	0.740	00401440
LU24	(110)13		60	8	90-138	0.375	0.825	0.746	GQ461442
	(770) 77777		50		450.004	0.000	0.000	0 75 4	00404454
LU25	$(11C)_{22}$	F: TCTACAGAGTTCAATTCCCGTAA	58	6	153-204	0.000	0.833	0.754	GQ461454
	1(110)7	R: GTIGGACCTTACAAGACTCACTG		_					
LU26	(TTC) ₁₁	F: CCTGCAGGAAAAAGTGAAGC	60	2	120-135	0.125	0.525	0.371	GQ461455
		R: CIGCAAACAGCACAACIICC							
LU27	(TTC) ₄ T(TTC) ₁	F: GTTTGAGAAGAGGGCATCCA	60	4	158-167	0.125	0.442	0.387	GQ461456
	7	R: GTTGGGGTGAAGAGGAACAA							
LU28	(TTC) ₁₀	F: GCTGTTAGCACTCAGCAGCA	60	2	134-137	0.000	0.233	0.195	GQ461459
		R: CACGTTGACCAACAAAACCA							
LU29	(ATC) ₅	F: GGGCAGTGATTGATTGGTTT	60	2	118-124	0.250	0.400	0.305	GQ461501
		R: GGCGGCAATTGCTACATT							
LU30	(ATC) ₅ (ATC)	F: TCTTGACCATCAGCATCACC	60	3	162-168	1.000	0.592	0.456	GQ461503
	6	R: GAGGCACAGGGAAACTAACG							
LU31	(ATC) ₁₂	F: TCTTTGTTTGGTGCCAAAGTT	58	4	97-112	0.250	0.592	0.510	GQ461514
		R: TTCATGATCTCACCTAACCTGA							
LU32	(ATC) ₁₂	F: ACGCGTAAACTTTCCGTTTC	60	3	144-150	0.000	0.633	0.511	GQ461515
		R: ATAATGTCGGCTGCTTCTGC							
LU33	(ATC) ₄	F: TTCTCCATCATCTCACATCCA	59	3	150-174	0.125	0.492	0.398	GQ461518
		R: CCAAATCAGAATGTGCGTGT							
LU34	(ATC) ₅	F: GGAAGAATTGGAAGAGGAAGG	60	2	128-134	0.375	0.325	0.258	GQ461524
		R: CCTTCTCCCATGATCAAACAA							
LU35	(ATC) ₅	F: CCAACGGATCATCCTCTAGC	61	2	143-149	0.000	0.233	0.195	GQ461532
		R: GGACAGAAAGGGGAAAGGAA							
LU36	(ATC) ₁₄	F: GAAGTTCGTGGGAGGAGTTG	60	2	127-130	0.000	0.400	0.305	GQ461534
		R: CCACATGAACCGCTGTAGAA							
LU37	(ATC) ₉	F: AGTACCCAACGCCCAGATTT	60	3	95-131	0.500	0.567	0.468	GQ461554
		R: AACAGCAGTCCTGGCAGTCT							
LU38	(ATC) ₉	F: GATCTTGTTGCCTGGGAAAG	60	2	80-104	0.125	0.325	0.258	GQ461564
		R: TTCGTTTGCAATACGTCAGC							
Mean				3.395		0.257	0.516	0.429	

Ta: Annealing temperature; k: number of allele; PIC: polymorphic information content; Ho: observed heterozygosity; HE: expected heterozygosity

previously reported 35 dinucleotide genomic SSR markers had an average of 3.457 alleles per locus (Deng et al., 2010). Roose-Amsaleg et al. (2006) reported that 28 genomic SSR markers exhibited an average of 3.32 when assessed on 93 flax cultivars, while the average alleles per locus of EST-SSR markers was only 2.3 based on 248 EST-SSRs assessed on 23 flax genotypes

(Cloutier et al., 2009). These may indicate that genomic SSR markers had higher alleles per locus than EST-SSR markers.

The PIC values ranged from 0.110 to 0.881 over 38 loci, with a mean value of 0.429 and 13 loci having PIC greater than 0.5. The observed heterozygosities for individual loci varied from 0.000 to 1.000 with an average

of 0.257 per locus. The expected heterozygosities varied from 0.125 to 0.950 with an average of 0.516. The relationship between the degree of polymorphism and the number of repeats has been reported in some species (Saghai et al., 1994; Fisher et al., 1998). But no clear correlation was found in flax from our data.

In conclusion, we constructed two microsatellite-enriched flax genomic libraries for trinucleotide units repeats. A total of 206 new microsatellite-containing sequences were identified and classified, and 38 polymorphic microsatellite markers were developed. These markers can potentially be used in flax for germplasm identification, genetic diversity study, and particularly in genetic mapping and breeding for marker assisted selection.

ACKNOWLEDGEMENTS

This work was funded by National Special Funds for Public Benefit Research of China (No. nyhyzx07-018-03) and the earmarked fund for Modern Agro-industry Technology Research System (No.nycytx-19-E14).

REFERENCES

- Adugna W, Labuschagne MT, Viljoen CD (2006). The use of morphological and AFLP markers in diversity analysis of linseed. Biodivers Conserv. 15: 3193-3205.
- Carol AR, Arcangeli L, Block AB (2003). Development of 11 mircrosatellite loci for population studies in the swordfish Xiphias gladius (Teleostei: *Scombridae*). Mol. Ecol. Notes, 3: 147-149.
- Cloutier S, Niu ZX, Datla R, Duguid S (2009). Development and analysis of EST-SSRs for flax (*Linum usitatissimum* L.). Theor. Appl. Genet. 119: 53-63.
- Cordeiro GM, Pan YB, Henry RJ (2003). Sugarcane microsatellites for the assessment of genetic diversity in sugarcane germplasm. Plant Sci. 165: 181-189.
- Deng X, Chen XB, Long SH, Wang XC, GAO Y, He DF, Wang J, Wang YF (2008). Microsatellite Marker Enrichment with Magnetic Beads in Flax. Acta. Agron. Sin. 12: 2099-2105.
- Deng X, Long ŠH, He DF, Li X, Wang YF, Liu J, Chen XB (2010). Development and characterization of polymorphic microsatellite markers in *Linum usitatissmum*. J. Plant Res. 123: 119-123.

- Fisher PJ, Richardson TE, Gardner RC (1998). Characteristics of single and multi-copy microsatellites from Pinus radiate. Theor. Appl. Genet. 96: 969-979.
- Fu YB, Peterson G, Diederichsen A, Richards KW (2002). RAPD analysis of genetic relationships of seven flax species in the genus *Linum L*. Genet. Resour. Crop. Environ. 49: 253-259.
- Fu YB, Rowland GG, Duguid SD, Richards KW (2003a). RAPD analysis of 54 North American flax cultivars. Crop Sci. 43: 1510-1515.
- Fu YB, Guerin S, Peterson GW, Diederichsen A, Rowland GG, Richards KW (2003b). RAPD analysis of genetic variability of regenerated seeds in the Canadian flax cultivar CDC Normandy. Seed Sci. Technol. 31: 207-211.
- Kalinowski ST, Taper ML, Marshall TC (2007). Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. Mol. Ecol. 16: 1099-1006.
- Li Q, Wan JM (2005). SSRHunter: development of a local searching software for SSR Sites. Hereditas, 27: 808-810 (in Chinese).
- Lunt DH, Hutchinson WF, Carvalho GR (1999). An efficient method for PCR-based isolation of microsatellite arrays (PIMA). Mol. Ecol. 8: 891-892.
- McBreen K, Lockhart PJ, McLenachan PA, Scheele S, Robertson AW (2003). The use of molecular techniques to resolve relationships among traditional weaving cultivars of Phormium. N Zeal. J. Bot. 41(2): 301-310.
- Millam S, Obert B, Pretova A (2005). Plant cell and biotechnology studies in *Linum usitatissimum* (a review). Plant Cell Tissue Organ Cult. 82: 93-103.
- Oh TJ, Gorman M, Cullis CA (2000). RFLP and RAPD mapping in flax (*Linum usitatissimum*). Theor. Appl. Genet. 101: 590-593.
- Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol. Breed. 2: 225-238.
- Roose Amsaleg C, Cariou Pham E, Vautrin D, Tavernier R, Solignac M (2006). Polymorphic microsatellite loci in *Linum usitatissimum*. Mol. Ecol. Notes, 6: 796-799.
- Rozen S, Skaletsky HJ (2000). Primer 3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds). Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 365-386.
- Saghai Maroof MA, Biyashev RM, Yang GP, Zhang Q, Allard RW (1994). Extraordinary polymorphic microsatellite DNA in barley: Species diversity, chromosomal locations and population dynamics. Proc. Natl. Acad. Sci. USA, 91: 5466-5470.
- Spielmeyer W, Green AG, Bittisnich D, Mendham N, Lagudah ES (1998). Identification of quantitative trait loci contributing to Fusarium wilt resistance on an AFLP linkage map of flax (*Linum usitatissimum*). Theor. Appl. Genet. 97: 633-641.