Genic microsatellite markers from expressed sequence tags (ESTs) of developing oat seed

Peter Eckstein, Aaron Beattie, Brian Rossnagel, and Graham Scoles.
Department of Plant Sciences/Crop Development Centre, University of Saskatchewan
Saskatoon, SK, Canada, S7N 5A8, (306) 966-8349, peter.eckstein@usask.ca

Introduction

Microsatellite markers (simple-sequence repeat (SSR)) have been widely available for the genetic analysis of crops such as wheat, barley, and rice, and formed the backbone of genetic studies in these crops due to high polymorphism and ease of use. Despite the benefits of SSR markers, few are available for oat genomics, and even fewer have been placed on genetic maps, a pre-requisite for their efficient association with traits of interest. Cross-applicability of wheat, barley, rice, and sorghum SSR markers to oat has generally been poor. Though polymorphism levels are adequate (if not high), polymorphisms are often dominant in nature resulting from primer-template mis-match, as opposed to the co-dominant, multi-allelic polymorphisms generated at true microsatellite loci and from which SSR markers gain analytical power. A small number of SSRs specific to the oat genome are available, and more are in development (Anderson, 2008). In addition, 195 oat specific SSR markers have recently been derived from oat EST sequences (Becher, 2007).

We have generated an EST sequence database in cooperation with the Natural Products Genomics Resource (NAPGEN) initiative of PBI-NRC. Some 19,680 ESTs were isolated and sequenced yielding 8062 unigenes (2418 contigs, 5644 singletons). This database was "mined" for regions containing simple-sequence repeats, similar to the study by Becher (2007). The SSR markers characterized, plus a large pool of potential SSRs not yet analysed, contribute to the growing number of SSR markers developed specifically from oat.

Materials and Methods

In silico analysis

Some 7000 unigenes were mined for SSR containing regions using the Web-based "SSR Primer Discovery" tool (http://hornbill.cspp.latrobe.edu.au). To avoid sequence redundancy, only unigenes were analysed. Primers flanking repeated motifs were designed using "Primer 3" (integrated in SSR Primer Discovery) set at default values (Robinson et al, 2004).

Genotypes

Kanota, Ogle, TAM O-301, Marion, Terra, Mn841801-1, Noble-2, CDC SolFi, HiFi (parents of four mapping populations and a CDC breeding cross).

SSR amplification

PCR reactions consisted of 20 mM Tris-HCl, 50 mM KCl, 2 mM MgCl, 400 uM dNTP, each primer @ 200 nM, 1 unit of *Taq* polymerase, and 100 ng template DNA in a volume of 25 uL. Amplification cycles (35X) comprised of 94°C (45 sec.), 55°C - 60°C (45 sec.), 72°C (60 sec.).

Fragment analysis

Denaturing polyacrylamide gel electrophoresis, 6 % acrylamide in 1X TBE. Gels run at constant W (60) for 2.5 hours. Fragments visualized using silver nitrate.

Results and Discussion

At search parameters similar to Becher (2007), *in silico* analysis identified 301 EST sequences (4.3%) containing perfect and imperfect simple-sequence repeats of n ≥ 6 for di-, n ≥ 5 for tri-, n ≥ 4 for tetra-, and n ≥ 4 for penta-nucleotide repeat motifs. Two-thirds of all SSRs were tri-nucleotide, consistent with results from many other genic SSR isolations. Numbers and types of SSRs mined from the database are summarized in Table 1. Approximately half (53%) of SSR containing sequences were BLASTX annotated to genes of known function, 15% to unknown function, while 32% of sequences were unique.

Sixty-one SSR primer pairs (Table 2), chosen on the basis of repeat lengths ≥ 20 nucleotides irrespective of motif length, were initially tested on the cultivar Kanota. Forty-two (69%) primer pairs mediated the amplification of one to three DNA fragments, five (8%) amplified weakly/not at all, and 14 (23%) amplified multi-fragment banding patterns.

Fifty-six SSR markers were analysed against nine genotypes representing parents of four mapping populations and a CDC breeding cross. Markers amplifying multi-band patterns were either monomorphic, displayed only dominant polymorphisms, or were difficult to interpret. Of 42 remaining markers, 25 generated co-dominant length polymorphisms in at least one genotype (Table 2), while 19 markers were polymorphic in multiple genotypes and could potentially be mapped in more than one population. Table 3 lists SSR markers polymorphic per mapping population. Primer sequences and annealing temperatures for polymorphic SSR markers, and for markers of excellent quality (though monomorphic or of longer than expected length), are given in Table 4.

Many amplified fragments were considerably longer than expected, likely a function of intron regions between marker priming sites, and common result in the development of EST-derived SSR markers. Many primer sets amplified multiple loci, a common result in hexaploid oat. The yield of polymorphic markers at 41% was somewhat lower than in other studies, perhaps a result of selection for good quality co-dominant polymorphisms only.

Current indications of marker quality and levels of polymorphism among nine oat cultivars (for the initial 61 SSR primer pairs), suggest good potential for the development of additional markers from the remaining pool of 240 SSRs.

Table 1. Summary of simple-sequence repeats identified by *in silico* analysis of 7000 oat ESTs.

Repeat Type	Minimum # of repeats	# of SSRs "mined"	% of Total # of SSRs
DI-nucleotide	**≥ 6**	58	19.3
TRI-nucleotide	**≥ 5**	202	67.1
TETRA-nucleotide	**≥ 4**	39	12.9
PENTA-nucleotide	**≥ 4**	2	0.7
		301	

Table 2. Summary of 61 characterized SSR markers by motif type, including repeat lengths and levels of polymorphism

Repeat Type	Minimum # of repeats	# of SSRs Tested	# (%) Polymorphic
DI-nucleotide	**≥ 10**	10	4 (40%)
TRI-nucleotide	**≥ 7**	38	16 (42%)
TETRA-nucleotide	**≥ 5**	11	5 (45%)
PENTA-nucleotide	**≥ 4**	2	0 (0%)
		61	25 (41%)

Table 3. Polymorphic SSR markers by population. K/O = Kanota x Ogle, O/TAM = Ogle x TAM 0-301, M/T = Marion x Terra, Mn/N = Mn841801-1 x Noble-2, Sol/Hi = CDC SolFi x HiFi.

SSR Marker	Population				
	K/O	O/TAM	M/T	Mn/N	Sol/Hi
OEM04			√		
OEM06	√	√		√	√
OEM08	√	√	√	√	
OEM09	√	√			
OEM11	√	√	√	√	√
OEM12		√			
OEM14		√		√	√
OEM16	√	√	√		√
OEM19		√			
OEM21			√		
OEM24	√	√	√		
OEM26	√	√	√	√	
OEM27	√	√		√	√
OEM32				√	√
OEM33					√
OEM37		√	√		
OEM38	√	√			
OEM40	√	√		√	√
OEM44	√	√			
OEM45	√	√			
OEM46	√	√	√	√	
OEM49		√			
OEM52	√	√	√		
OEM53	√		√	√	√
OEM55	√	√		√	
Per Population	16	20	12	11	9

Table 4. PCR annealing temperatures, marker quality scores, repeat motifs, and primer sequences, for polymorphic SSR markers, and for high quality monomorphic SSR markers. Quality ratings are those of Stephenson et al. (1998); 1 = single high quality locus, 2 = single locus, 3 = single locus sometimes difficult to interpret, 4 = multiple loci, easy to score, but may be difficult to relate polymorphic loci.

SSR Marker	Tm (° C)	Quality	Repeat Motif/Length (CDC Dancer)	Left Primer	Right Primer
Polymorphic SSR Markers					
OEM04	57	1	$(CTG)_8$	ATCGCGCTGTACCTATAATC	GCTCCAAACAACACGTAAAT
OEM06	57	4	$(AGG)_8$	CACGCAGCACCACTACTAC	GGAGGCTGTTGTTGTTATTG
OEM08	57	1	$(CAC)_8$	GTGACTCCTCAACCACACA	CTTGATGTAGCGTCCGTAGT
OEM09	57	1	$(AAG)_{10}$	CAGGAAATAATATAACGCCCG	CTTCTCAGATTTGGAACTCG
OEM11	57	1	$(GA)_{11}$	AGAGAGGCAGAACAGAACAA	CTGTCCTTGACATCTCCTTC
OEM12	57	1	$(GA)_{11}$	TTCTTCTTCTTCTTGTTGGC	GCAAGTACTGTGTTGCAGAA
OEM14	57	1	$(CCGT)_6$	CGTCCTGTGTCTTTACTTCC	AGCACCAGATGGTCTACAAC
OEM16	57	4	$(GA)_{10}$	AGCAATTCAAAGCATCTCC	GTATCTGCCCTTCTTGAGG
OEM19	57	4	$(CTC)_7$	CACGCTATTGATTGCTGC	GGTATGATGATATTGACGGC
OEM21	57	5	$(GCG)_7$	CTCTGTTTGTTGAGTGGACA	GGTAAATGTGACTGGGAAGA
OEM24	57	4	$(GCC)_7$	GAAATCTCCTCCTCCTCTTC	TCTCTTGTTCCACGTAGTCA
OEM26	60	4	$(ACA)_7$	AGTATGACCCTAGCGAACAA	AACTCCGATGGTCCATTTAT
OEM27	57	2	$(CGG)_7$	CAGCTAAATGTACTGAGCCC	AGTTGGATCAGCAGGAAGA
OEM32	60	4	$(AAG)_4GAG(AAG)_3$	ACGATGATCACAAGAAGGAC	GGAGCTCAGTCGCTGTCG
OEM33	57	1	$(CAC)_6CA$	TTTCGTCGTATCTAAGGCAT	AGCAGAAGGTGATGGAGTAA
OEM37	60	4	$(CTT)_6CT$	AGACAAGAATTTCTCCCAT	GTGGAGATATGGTCGTTGTT
OEM38	57	1	$(GGA)_6GG$	CTGTCTCAAATGGATCCTG	TCCTAAGCTTTCTGTTCTCG
OEM40	57	3	$(AGCT)_6$	CCACAATTACGATGAGGAGT	TCAAGCACTATGCAGTTGTC
OEM44	57	3	$(TCGG)_6$	GGTAGTCGTCGTCCTCCT	CTAACTTCTGGTCCTTGACG
OEM45	57	3	$(CCTC)_6$	CTCCGGAAGTGAACGAAG	GCTAGCAGAAATGAGAGGAA
OEM46	55	4	$(CAG)_4CAA(CAG)_3$	AGATGCAGCAAACATCTTCT	CTGTTGTTCTGTTGCATTTG
OEM49	55	3	$(GT)_6CT(GT)_6$	TGGCGTGAGAGTATACATGA	AAGACCAGAGCATAGCAGAG
OEM52	60	3	$(GCC)_7$	GAAATCTCCTCCTCCTCTTC	TCTCTTGTTCCACGTAGTCA
OEM53	60	3	$(TGC)_6TG$	GTCAACCGCTAAGTTGAGTT	AAACAGCAGATGATTTGGAG
OEM55	60	3	$(AGCT)_6$	CCACAATTACGATGAGGAGT	TCAAGCACTATGCAGTTGTC
High Quality SSR Markers but Monomorphic in 9 Genotypes Tested					
OEM01	60	1	$(AC)_{12}$	CGTCTTCTTGATCTTTCCTG	TATGATAACCGTTCCCTGAC
OEM02	60	2	$(GA)_{10}$	CTTCGCTTCTCCGATTTG	ATAGTACACATCAGCCGTCC
OEM05	57	2	$(GTG)_6$	GGTAGAGACCGTAATGGTGA	TCTAGCAAAGTGTCCAGGTT
OEM15	57	1	$(CCG)_6$	CACTCCTCCAAACCCTTT	GATCTCGTCGATACGGAATA
OEM17	57	1	$(CGAT)_5CGAA$	ACACCGCCCGAAGCGAAAC	AAGATCAAGAACGCAGCTCTC
OEM25	57	1	$(GAA)_5GA(GAA)_3$	AACCCTAGCCAGGCTCTA	CTCGAAGTGCCCTGAGTTAC
OEM30	57	1	$(CTC)_7$	AACACGGAAGCAAGGGTA	TACATCTGGAGATGGCTCTT
OEM31	57	1	$(GCA)_7$	GTGAAGTCAATGCAGATCG	AGGATCTCCTCCCTGTACTC
OEM34	57	1	$(GGC)_7$	TATTCTGCTTTGCCCTCATA	TCTGTGGATATTTGGAAGG
OEM39	57	1	$(AGG)_7$	TGTAGTCCAGGAAGCACTCT	TCTCTTTGCTCAACTGCTAA
OEM47	57	1	$(TGG)_4AGCCGG(TGG)_3$	TAAGAGTGGAACCTGCTCAT	ACTGCCTTCACGATACACTT
OEM57	60	1	$(GT)_6CT(GT)_6$	TGGCGTGAGAGTATACATGA	AAGACCAGAGCATAGCAGAG
OEM59	60	3	$(GAT)_3GAA(GAT)_3$	GATTACCTTCTGTCTCGCAG	TCTCCTCTTCAGTCTCCTCA
OEM60	60	3	$(GAT)_3GAA(GAT)_3$	CAGCTGTACCTGTAGCATCA	TAACCATGTCAGTCTCGTCA

References

Anderson, J. 2008. Overview of oat microsatellite development to date. Plant and Animal Genome XVI. San Diego, CA, January 12-16, 2008.

Becher, R. 2007. EST-derived microsatellites as a rich source of molecular markers for oats. Plant Breeding 126(3), 274–278.

Robinson, A.J., Love, C.G., Batley, J.,, Barker, G., and D. Edwards. 2004. Simple sequence repeat marker loci discovery using SSR primer. Bioinformatics 20(9), 1475-1476.

Stephenson, P., Bryna, G., Kirby, G., Collins, A., Devos, K., Busso, C., and M. Gale. 1998. Fifty new microsatellite loci for the wheat genetic map. Theor. Appl. Genet. 97, 946-949.

Acknowledgements

This work was conducted and funded by the NAPGEN (Natural Products Genomics Resource) initiative of PBI-NRC. Additional funding was provided through Crop Development Centre oat variety pedigreed seed royalty funds.