Development and Characterization of 37 Novel EST-SSR Markers in Pisum sativum (Fabaceae)

Authors: Zhuang, Xiaofeng, McPhee, Kevin E., Coram, Tristan E., Peever, Tobin L., and Chilvers, Martin I.

Source: Applications in Plant Sciences, 1(1)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1200249

Applications in Plant Sciences 2013 1(1): 1200249

PRIMER NOTE

DEVELOPMENT AND CHARACTERIZATION OF 37 NOVEL EST-SSR MARKERS IN *PISUM SATIVUM* (FABACEAE)[1]

XIAOFENG ZHUANG[2], KEVIN E. MCPHEE[3], TRISTAN E. CORAM[4], TOBIN L. PEEVER[5], AND MARTIN I. CHILVERS[2,6]

[2]Department of Plant, Soil and Microbial Sciences, Michigan State University, 107 CIPS, East Lansing, Michigan 48824 USA; [3]Department of Plant Sciences, North Dakota State University, 370G Loftsgard Hall, Fargo, North Dakota 58108 USA; [4]Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, Indiana 46268 USA; and [5]Department of Plant Pathology, Washington State University, Pullman, Washington 99164 USA

- *Premise of the study:* Simple sequence repeat markers were developed based on expressed sequence tags (EST-SSR) and screened for polymorphism among 23 *Pisum sativum* individuals to assist development and refinement of pea linkage maps. In particular, the SSR markers were developed to assist in mapping of white mold disease resistance quantitative trait loci.
- *Methods and Results:* Primer pairs were designed for 46 SSRs identified in EST contiguous sequences assembled from a 454 pyrosequenced transcriptome of the pea cultivar, 'LIFTER'. Thirty-seven SSR markers amplified PCR products, of which 11 (30%) SSR markers produced polymorphism in 23 individuals, including parents of recombinant inbred lines, with two to four alleles. The observed and expected heterozygosities ranged from 0 to 0.43 and from 0.31 to 0.83, respectively.
- *Conclusions:* These EST-SSR markers for pea will be useful for refinement of pea linkage maps, and will likely be useful for comparative mapping of pea and as tools for marker-based pea breeding.

Key words: EST-SSR; Fabaceae; microsatellite; *Pisum sativum*; *Sclerotinia sclerotiorum*; transcriptome.

Pea (*Pisum sativum* L.) is one of the most important legumes grown and consumed worldwide. White mold caused by the fungal pathogen *Sclerotinia sclerotiorum* (Lib.) de Bary is a significant yield-limiting disease of pea in most areas that pea is cultivated. Despite the agricultural importance of pea, pea breeding is constrained by a large genome size (~4300 Mb), lack of genomic resources, and rich repetitive DNA (estimated at 75–97% of the pea genome) (Macas et al., 2007). Molecular markers have great potential to speed up the process of developing improved cultivars. Although several hundred simple sequence repeat (SSR) markers have been identified (Burstin et al., 2001; Loridon et al., 2005; Gong et al., 2010), additional SSR markers with polymorphism are needed, particularly for the development of linkage maps for use in white mold–resistance mapping studies.

With the development of next-generation sequencing technologies, large amounts of expressed sequence tags (ESTs) have been generated for model species as well as economically important nonmodel plants. These ESTs offer an opportunity to discover novel genes and have also provided a resource to develop markers (Davey et al., 2011). Recently, we sequenced the transcriptome of pea infected by *S. sclerotiorum* using next-generation sequencing to understand this host–pathogen interaction. The transcriptome sequences from pea contain abundant SSRs, which we have used in this study to develop SSR markers. The SSR markers were screened against 23 pea cultivars and plant introductions (PIs), including parents of four recombinant inbred line (RIL) populations (Lifter and PI240515; Medora and PI169603; Bohatyr and Shawnee; Melrose and Radley) for white mold–resistance mapping studies. These new markers will be very useful for linkage mapping studies.

[1]Manuscript received 22 May 2012; revision accepted 15 July 2012.

Funding for this study was provided by the National Sclerotinia Initiative (USDA-ARS Specific Cooperative Agreement #58-5442-9-239).

[6]Author for correspondence: chilvers@msu.edu

doi:10.3732/apps.1200249

METHODS AND RESULTS

LIFTER, a cultivar susceptible to *S. sclerotiorum* (McPhee and Muehlbauer, 2002), was inoculated with *S. sclerotiorum* isolate WMA-1 (≡ATCC MYA-4521) on the stem between the fourth and fifth detectable nodes. Seventy-two hours after inoculation, total RNA was extracted from 18 infected plants by cutting a 1 cm piece of pea stem containing the advancing lesion front toward the base of the plant using the TRIzol Plus RNA Purification Kit (Invitrogen, Carlsbad, California, USA). Messenger RNA was purified from the total RNA with the Oligotex mRNA Mini Kit using the mRNA Spin-Column Protocol (QIAGEN, Valencia, California, USA) and converted into a normalized cDNA pool with the services of Evrogen (http://www.evrogen.com). Transcriptome sequencing of pea infected by *S. sclerotiorum* was conducted on a full plate of the Roche 454 GS FLX sequencer (454 Life Sciences, Branford, Connecticut, USA) at Washington State University. In total, 128 720 high-quality reads with an average length of 215 nucleotides were obtained and assembled into 10 158 contiguous sequences (contigs) with the program ABySS (Simpson et al., 2009). Pea and *S. sclerotiorum* contigs were parsed with a tBLASTx method (Zhuang et al., 2012) against publicly available, closely related plant and fungal genome databases. The fungal genome database consisted of *S. sclerotiorum* (strain1980) and six closely related fungal (Ascomycete) species (*Botrytis cinerea* Pers., *Chaetomium globosum* Kunze, *Fusarium graminearum* Schwabe, *Magnaporthe grisea* (T. T. Hebert) M. E. Barr, *Neurospora crassa* Shear & B. O. Dodge, and *Verticillium dahlia* Kleb.), and the plant genome database consisted of three sequenced legume (Fabaceae) genomes (*Glycine max* (L.) Merr., *Lotus japonicus* (Regel) K. Larsen, and *Medicago truncatula* Gaertn.). After parsing, 10 158 contigs were separated into 6299 pea ESTs, 2780 *S. sclerotiorum* ESTs, and 1079 unassigned ESTs. Among the pea ESTs, 118 potential

TABLE 1. Characteristics of 11 polymorphic and 26 monomorphic *Pisum sativum* EST-SSR markers.

Locus	Primer sequences (5′–3′)	Repeat motif	Size (bp)	T_a (°C)	GenBank accession no.	Putative function [organism]	*E*-value
Psat61*	F: CCGGTTCGGTTTCCGGTTGAGG	$(GGGTTC)_4$	81	60	JR344273	unknown	
	R: ACGGACTCCAGCCAGCACCA						
Psat900*	F: GCTGATCCCATTCCAACCACAGGC	$(TTG)_5$	135	58	JR344282	chalcone reductase [*Medicago sativa*]	7.44E-12
	R: ACAACCTTACCTTAAACCTTCTCAACC						
Psat921*	F: TCAACTCTCAACAGGCGCTGC	$(GCT)_6$	248	58	JR344284	unknown	
	R: TGTCACAACGACCCTGCAAGC						
Psat5404*	F: ACTTCACATTGCACTCTTTCTTCAC	$(CT)_5$	103	56	JR344267	unknown	
	R: TGAATCTCCCATATCTCAACTCAAGTG						
Psat5545*	F: TCCCATGGAACAAGCTCATCATCC	$(TCA)_7$	123	58	JR344268	predicted protein [*Medicago truncatula*]	3.98E-05
	R: TGGGTTCAGTGAGGAACAGGT						
Psat5571*	F: AGGAGCGGCTGAAGAAAGAGT	$(AG)_6$	135	58	JR344269	predicted protein [*Glycine max*]	5.32E-26
	R: CACCGCTGTAGAGGGCGTGA						
Psat7112*	F: TGATGATGTGCTGATTATTGTTCTGGT	$(TTA)_5$	170	58	JR344274	unknown	
	R: ACAGTCACAGAAAGTGTCTACAGCA						
Psat7598*	F: ACTACAGGAGTTGAATTTGCGGA	$(GAT)_6$	209	54	JR344275	basic helix-loop-helix protein bhlh5 [*Lotus japonicus*]	3.43E-04
	R: CAACATCAACAAGAACAAGAACACG						
Psat7818*	F: TTGAGGTTGTTGTTGTTGTTGCTGT	$(GTT)_5$	80	58	JR344277	predicted protein [*Ricinus communis*]	1.82E-05
	R: AAACAAAGGAAGTTTGGGCAGC						
Psat9662*	F: AGTGAAGCGAGTGGAAGATACGA	$(GAAATC)_5$	171	58	JR344287	fiber protein fb11 [*Camellia sinensis*]	5.59E-15
	R: GGCCAAAGCCGGCGATGAGA						
Psat10014*	F: ATATCGCCACGACGCAAAGC	$(GT)_6$	124	58	JR344254	unknown	
	R: TCTTACATGACAAAGCCAACACAAG						
Psat368	F: ACATTCCTCCGGCGTAGCTGA	$(AATCGG)_4$	81	58	JR344261	DNA repair protein Rad23-1 [*Ricinus communis*]	1.28E-27
	R: ACAGTGAGCTTCATGACTACTCGGC						
Psat373	F: CCTGGTGATGCTCCTCAGGCA	$(GAA)_5$	155	58	JR344262	transcription elongation factor family protein [*Arabidopsis thaliana*]	3.31E-39
	R: TCAGCTGTAATCTCAAGCTCAGCCA						
Psat589	F: TGTGTAGCATCATCAGCGGAGC	$(GA)_5$	162	58	JR344271	chitinase [*Cicer arietinum*]	1.54E-96
	R: CCCGCAACTAAACCTTGCTGGC						
Psat1176	F: GCCTATTTGTACTATTCCACCACCTG	$(TA)_5$	179	56	JR344256	unknown	
	R: ACGGATGAATAAGTGACATTACAGTGA						
Psat1764	F: TCAGGGTCGGTGAGGCTTCGT	$(GA)_6$	152	58	JR344257	predicted protein [*Medicago truncatula*]	6.40E-11
	R: TCAGTGAAGAACATGGCACCAA						
Psat2045	F: GAAGCGGCGACGATGGCGTA	$(TTG)_5$	224	58	JR344258	mannose-P-dolichol utilization defect 1 protein [*Arabidopsis thaliana*]	5.18E-21
	R: AGTTCAGTTTGAGTAAACATTGACGG						
Psat2885	F: AGACGGAGGAGACGTGGAGGA	$(GA)_5$	139	58	JR344259	unknown	
	R: CACCACCACCAACGCCGTCA						
Psat3352	F: TGATTGGGATCGACTTCGACGG	$(GA)_5$	100	58	JR344260	40S ribosomal protein S5 [*Cicer arietinum*]	2.99E-08
	R: AGAGCATTTGAAGTGTTTACGGCTGC						
Psat4097	F: GCCAAACATGCCAACAACAATCCT	$(TTA)_6$	152	58	JR344263	unknown	
	R: TCACTGAGCCACCGCCAACG						
Psat4741	F: CCACCACTTCAACCCTCTCAACGA	$(GTG)_5$	123	56	JR344264	Phloem-specific protein [*Pisum sativum*]	1.34E-24
	R: TCGACCGCTACCCAAACGCTG						
Psat4773	F: ACAGCTCCTGGCACAGCTCTT	$(TA)_5$	178	58	JR344265	predicted protein [*Glycine max*]	1.57E-25
	R: CCCAATTGCTTATGTCTGCTGCCT						
Psat5398	F: TCACCAATTCGCCCTCTCTCCA	$(CT)_5$	131	58	JR344266	unknown	
	R: CGCAAGGTTCCAGATTCTTCGAGGT						
Psat5852	F: TGCCAACCAGGTAGAGTCTCA	$(CTT)_5$	151	56	JR344270	predicted protein [*Medicago truncatula*]	8.04E-30
	R: AGTCGAATCTTGTTCCTTCTTCTTTGA						
Psat6026	F: TGTGCTTCTTGTGGCTGGTGA	$(GA)_5$	134	58	JR344272	predicted protein [*Glycine max*]	1.40E-34
	R: GTCCCTCGCGACGACACCAA						
Psat7675	F: TCACGTCGCTTCGTTTCATCCC	$(TGA)_6$	183	56	JR344276	small heat shock protein 1 [*Prunus salicina*]	2.94E-08
	R: ACCACCCATCACACCAAACCCA						
Psat7820	F: CCGGAGCGGAGGCGAAGAGA	$(TTC)_5$	159	56	JR344278	predicted protein [*Medicago truncatula*]	6.37E-35
	R: GGGACGCAGTAATCAACCAGA						
Psat7825	F: CCAGACACAGATCCTCAACAACTCCG	$(CT)_5$	94	60	JR344279	intracellular chloride channel [*Medicago truncatula*]	5.76E-28
	R: GCGGCGCACTTTCGTAGCAG						
Psat8001	F: TCTCCTCACAACTCAACTGTTACC	$(TTC)_7$	141	58	JR344280	unknown	
	R: TGGTGGTGAGACCGAGTGAGA						
Psat8487	F: TGTTTCCAGAAGGTTATGGCCC	$(GA)_5$	155	54	JR344281	HXXXD-type acyl-transferase-like protein [*Glycine max*]	8.16E-22
	R: AGATTCTTCGTTAGCCTTTGCTTTGA						
Psat9191	F: TGCAAACTTCAATGAGAGATCTGAAAG	$(TA)_5$	158	56	JR344283	predicted protein [*Medicago truncatula*]	2.03E-04
	R: TGCATTGGAGATGCCAAATCTGACT						
Psat9319	F: GCAGCACCACCACTCGCAGG	$(CCA)_5$	105	56	JR344285	unknown	
	R: AGCTGAGGTGATTGCTTCTGGT						
Psat9501	F: GCTTGCCTTTTGATTTTCCACGTCA	$(CTT)_5$	196	56	JR344286	proteinase inhibitor I4, serpin [*Medicago truncatula*]	1.77E-70
	R: TCATCGTGCGGTTGCACTTGT						

TABLE 1. Continued.

Locus	Primer sequences (5′–3′)	Repeat motif	Size (bp)	T_a (°C)	GenBank accession no.	Putative function [organism]	*E*-value
Psat9677	F: TGCAACAACTACGGATCACCAGC	$(TTGTA)_4$	189	58	JR344288	unknown	
	R: GCTGAACCAGATACACAAAGTTGAGC						
Psat9736	F: GGTCCTCCTCCAGGTTATGACCCTC	$(GAAA)_4$	108	56	JR344289	ensangp00000004563 related [*Medicago truncatula*]	6.40E-19
	R: AGTTTCCTTTACCTGAAGTCGTTTCT						
Psat9907	F: GACGGAACCGCCGTCCAACA	$(GAC)_5$	162	58	JR344290	predicted protein [*Glycine max*]	3.99E-05
	R: ACCACCTTGAGCGGCGTCAT						
Psat10084	F: TGCGGAGAAAGCGCTGCTGG	$(AAG)_5$	110	56	JR344255	unknown	
	R: ACGCAACCTTCTTCTTCTTCTTTCT						

Note: T_a = annealing temperature.
* Polymorphic EST-SSR markers.

SSRs, with more than five repeat units or a minimum repeat size of 20 nucleotides, were identified in 112 contigs of pea with the program SSRIT (Temnykh et al., 2001; Appendix S1). Of these 118 SSRs, trinucleotide repeats represented the largest fraction (50%) followed by dinucleotide (39.8%) SSRs. Two tetranucleotide, three pentanucleotide, and seven hexanucleotide SSRs were also identified in this pool. It was possible to design primers to the SSR flanking regions of 46 of the 118 SSRs using Primer3 (Rozen and Skaletsky, 2000; Table 1) with default parameters.

The SSR markers were tested against 23 individual pea cultivars (Appendix 1), including parents of four pea RIL mapping populations (Lifter and PI240515; Medora and PI169603; Bohatyr and Shawnee; Melrose and Radley), which are being used to map quantitative trait loci for resistance to white mold. Genomic DNA from each individual was extracted from leaves using the DNeasy Plant Mini Kit (QIAGEN). PCR contained 4 µL of 5× GoTaq PCR Buffer (Promega Corporation, Madison, Wisconsin, USA), 200 µM each dNTP, 2.5 µM each primer, 0.4 U of GoTaq polymerase, and ~50 ng of DNA template in a final volume of 20 µL. PCR were held at 94°C for 2 min; followed by 35 cycles of 94°C for 30 s, 55–60°C for 30 s, and 72°C for 1 min; with a final extension at 72°C for 10 min. The PCR products were separated in 10% polyacrylamide gels run in a Mega-Gel high-throughput electrophoresis system for 2.5 h at 300 V (C.B.S. Scientific, San Diego, California, USA). SSR bands were visualized with ethidium bromide, which was added to the running buffer. SSR band size was calculated by comparison with a 25-bp DNA ladder (Invitrogen). PCR products with expected sizes were successfully amplified for 37 primer sets, among which 11 showed clear polymorphisms with two to four alleles (Table 2). Observed heterozygosity and expected heterozygosity were calculated using POPGENE (version 1.32; Yeh and Boyle, 1997), and ranged from 0 to 0.43 and from 0.31 to 0.83, respectively. Ten of 11 markers (except Psat7598) were polymorphic in parents of at least one RIL population for white mold–resistance mapping studies. To determine if there was any redundancy between the SSRs described in this study and those previously published, all 37 ESTs were executed with BLASTn against *P. sativum* EST databases in the National Center for Biotechnology Information (taxid: 3888) with a cutoff parameter of $1e^{-20}$. BLASTn results show that only one EST (Psat4741) matched to a previously described but unpublished SSR marker; all other 36 ESTs including all 11 polymorphic SSRs were found to be unique to this study.

TABLE 2. Results of initial primer screening in 23 *Pisum sativum* individuals.

Locus	*A*	H_o	H_e
Psat61	3	0.0000	0.6367
Psat900	2	0.4348	0.513
Psat921	2	0.0000	0.513
Psat5404	2	0.0000	0.5362
Psat5545	2	0.0000	0.7063
Psat5571	4	0.3043	0.3092
Psat7112	3	0.0000	0.6947
Psat7598	2	0.0909	0.8309
Psat7818	2	0.0000	0.6
Psat9662	3	0.1739	0.487
Psat10014	2	0.1304	0.7362

Note: *A* = number of alleles; H_e = expected heterozygosity; H_o = observed heterozygosity.

CONCLUSIONS

In this study we demonstrate that next-generation sequencing is an effective tool to rapidly develop EST-derived SSR markers. We identified 37 *P. sativum* EST-SSRs, with 11 being polymorphic in 23 *P. sativum* individuals. These novel EST-SSR markers will be valuable tools for marker-assisted breeding, development of pea linkage maps, and comparative mapping of pea.

LITERATURE CITED

BURSTIN, J., G. DENIOT, J. POTIER, C. WEINACHTER, G. AUBERT, AND A. BARRANGER. 2001. Microsatellite polymorphism in *Pisum sativum*. *Plant Breeding* 120: 311–317.

DAVEY, J. W., P. A. HOHENLOHE, P. D. ETTER, J. Q. BOONE, J. M. CATCHEN, AND M. L. BLAXTER. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews. Genetics* 12: 499–510.

GONG, Y. M., S. C. XU, W. H. MAO, Q. Z. HU, G. W. ZHANG, J. DING, AND Y. D. LI. 2010. Developing new SSR markers from ESTs of pea (*Pisum sativum* L.). *Journal of Zhejiang University, Science B* 11: 702–707.

LORIDON, K., K. MCPHEE, J. MORIN, P. DUBREUIL, M. PILET-NAYEL, G. AUBERT, C. RAMEAU, ET AL. 2005. Microsatellite marker polymorphism and mapping in pea (*Pisum sativum* L.). *Theoretical and Applied Genetics* 111: 1022–1031.

MACAS, J., P. NEUMANN, AND A. NAVRÁTILOVÁ. 2007. Repetitive DNA in the pea (*Pisum sativum* L.) genome: Comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. *BMC Genomics* 8: 427.

MCPHEE, K. E., AND F. J. MUEHLBAUER. 2002. Registration of 'LIFTER' green dry pea. *Crop Science* 42: 1377–1378.

ROZEN, S., AND H. SKALETSKY. 2000. Primer3 on the WWW for general users and for biologist programmers. *In* S. Misener and S. A. Krawetz [eds.], Methods in molecular biology, vol. 132: Bioinformatics methods and protocols, 365–386. Humana Press, Totowa, New Jersey, USA.

SIMPSON, J. T., K. WONG, S. D. JACKMAN, J. E. SCHEIN, S. J. JONES, AND I. BIROL. 2009. ABySS: A parallel assembler for short read sequence data. *Genome Research* 19: 1117–1123.

TEMNYKH, S., G. DECLERCK, A. LUKASHOVA, L. LIPOVICH, S. CARTINHOUR, AND S. MCCOUCH. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. *Genome Research* 11: 1441–1452.

YEH, F. C., AND T. J. BOYLE. 1997. Population genetic analysis of co-dominant and dominant markers and quantitative traits. *Belgian Journal of Botany* 129: 157.

ZHUANG, X., K. E. MCPHEE, T. E. CORAM, T. L. PEEVER, AND M. I. CHILVERS. 2012. Rapid transcriptome characterization and parsing of sequences in a non-model host-pathogen interaction; pea-*Sclerotinia sclerotiorum*. *BMC Genomics* 13: 668.

Appendix 1. Information on 23 *Pisum sativum* germplasm lines used in this study; germplasm lines were obtained from the collection of Dr. Kevin McPhee's pea breeding program.[a] Information presented: country of origin, name, registration number.

USA: LIFTER, PI 628276*. MEDORA, N/A. MELROSE, PI 618628*. NDP080111, N/A. PI160936, N/A. PI240515, N/A. PS03101269, N/A. PS05ND0327, N/A. PS05ND0330, N/A. PS05ND0434, N/A. PS07ND0190, N/A. SHAWNEE, PI 619079*. SPECTER, PI 641005. STIRLING, PI 634571. WINDHAM, PI 647868.

Canada: AGASSIZ, 6093. CDC GOLDEN, 5602. CDC STRIKER, 5550. DS ADMIRAL, 5166. MAJORET, N/A.

Europe: BOHATYR, N/A. COOPER, N/A.

Note: N/A = not available.

[a]Pea cultivars, Plant Introductions (PIs), or breeding material not located in the Germplasm Resources Information Network (GRIN) are available from Dr. Kevin McPhee upon request. Voucher specimens have not been deposited due to their availability either within GRIN or the pea breeding community; additionally, some germplasm lines are the property of Dr. McPhee and North Dakota State University.

*Pea cultivars available from GRIN (http://www.ars-grin.gov/).