Polymorphism of PCR-based markers targeting exons, introns, promoter regions, and SSRs in maize and introns and repeat sequences in oat[1]

J.B. Holland, S.J. Helland, N. Sharopova, and D.C. Rhyne

Abstract: Sequence databases could be efficiently exploited for development of DNA markers if it were known which gene regions reveal the most polymorphism when amplified by PCR. We developed PCR primer pairs that target specific regions of previously sequenced genes from *Avena* and *Zea* species. Primers were targeted to amplify 40 introns, 24 exons, and 23 promoter regions within 54 maize genes. We surveyed 48 maize inbred lines (previously assayed for simple-sequence repeat (SSR) polymorphism) for amplification-product polymorphism. We also developed primers to target 14 SSRs and 12 introns within 18 *Avena* genes, and surveyed 22 hexaploid oat cultivars and 2 diploid *Avena* species for amplification-product polymorphism. In maize, 67% of promoter markers, 58% of intron markers, and 13% of exon markers exhibited amplification-product polymorphisms. Among polymorphic primer pairs in maize, genotype diversity was highest for SSR markers (0.60) followed by intron markers (0.46), exon markers (0.42), and promoter markers (0.28). Among all *Avena* genotypes, 64% of SSR markers and 58% of intron markers revealed polymorphisms, but among the cultivars only, 21% of SSR markers and 50% of intron markers were polymorphic. Polymorphic-sequence-tagged sites for plant-breeding applications can be created easily by targeting noncoding gene regions.

Key words: *Avena*, *Zea*, genetic diversity, DNA sequence.

Résumé : Les bases de données de séquence pourraient être exploitées pour le développement de marqueurs moléculaires si les régions révélant le plus de polymorphisme suite à une amplification PCR étaient connues. Les auteurs ont synthétisé des paires d'amorces qui amplifient des régions particulières au sein de gènes déjà séquencés chez les genres *Avena* et *Zea*. Les amorces ciblaient 40 introns, 24 exons, et 23 régions promotrices provenant de 54 gènes du maïs. Les auteurs ont examiné 48 lignées fixées de maïs (caractérisées précédemment à l'aide de microsatellites) pour du polymorphisme au niveau des amplicons. Les auteurs ont également développé des amorces afin de cibler 14 microsatellites et 12 introns situés dans 18 gènes du genre *Avena* pour ensuite examiner le niveau de polymorphisme parmi 22 cultivars d'avoine hexaploïdes et deux espèces diploïdes d'*Avena*. Chez le maïs, 67 % des marqueurs situés dans des promoteurs, 58 % des marqueurs situés dans des introns, et 13 % des marqueurs situés dans des exons ont montré du polymorphisme. Parmi les amorces révélant du polymorphisme chez le maïs, la diversité génotypique était la plus grande pour les microsatellites (0,60), suivie des marqueurs situés dans les introns (0,46), dans les exons (0,42) et finalement dans les promoteurs (0,28). Parmi tous les génotypes du genre *Avena*, 64 % des microsatellites et 58 % des marqueurs situés dans des introns étaient polymorphes. Des marqueurs STS polymorphes peuvent ainsi être développés aisément pour diverses applications en amélioration génétique en ciblant des régions non-codantes.

Mots clés : *Avena*, *Zea*, diversité génétique, séquence d'ADN.

[Traduit par la Rédaction]

Introduction

Public databases of plant gene and protein sequences represent a valuable and growing resource for plant genetics and breeding. For example, simple-sequence repeats (SSRs) can often be detected in gene sequences simply by searching for diagnostic repeat sequences. Each SSR is a tandem repeat of one or more short, simple sequences of two to six nucleotides. Polymorphism is detected using oligonucleotide primers complementary to conserved sequences flanking re-

Received March 13, 2001. Accepted August 16, 2001. Published on the NRC Research Press Web site at http://genome.nrc.ca on November 2, 2001.

Corresponding Editor: G.J. Scoles.

J.B. Holland[2] and D.C. Rhyne. USDA–ARS Plant Science Research Unit, Department of Crop Science, Box 7620, North Carolina State University, Raleigh, NC 27695, U.S.A.
S.J. Helland. Department of Plant Pathology, Iowa State University, Ames, IA 50011, U.S.A.
N. Sharopova. Department of Agronomy, University of Missouri—Columbia, Columbia, MO 65211, U.S.A.

[1]Mention of a trademark or proprietary product does not constitute a guarantee or warranty by USDA–ARS and does not imply approval over other products that may also be suitable.
[2]Corresponding author (e-mail: james_holland@ncsu.edu).

 DOI: 10.1139/gen-44-6-1065

peats in a polymerase chain reaction (PCR). Gel electrophoresis can then be used to separate PCR amplification products according to size, allowing detection of differences in the number of repeats targeted and amplified. SSRs are favored for plant-breeding and genetics applications because they are abundant in plant genomes, highly polymorphic within species, relatively rapid and inexpensive to assay, and can be used to identify specific chromosomal regions consistently across populations (Chen et al. 1997; Chin et al. 1996; Senior et al. 1998; Smith et al. 1997; Taramino and Tingey 1996). SSR development in oat (*Avena sativa* L.) lags behind that in other major cereal crops, with only 16 SSRs exhibiting polymorphism between cultivated genotypes reported to date (Li et al. 2000).

SSRs can be identified empirically by screening DNA libraries for repeat motifs via hybridization and sequencing candidate clones (Li et al. 2000; Taramino and Tingey 1996), applying SSR primers from related species (Li et al. 2000; Westman and Kresovich 1998) or mining sequence databases (Chin et al. 1996; Senior and Heun 1993; Wang et al. 1994). GenBank, a database supported by the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/), is the major public source of plant gene sequence information. As of July 2001, GenBank contained 111 319 total sequences and 214 intron-containing sequences from *Zea mays* L. and 186 854 total sequences and 1020 intron-containing sequences from *Arabidopsis thaliana*. Resources for *Avena* species, however, are significantly smaller. The database included 767 cDNA, genomic, chloroplast, intron, spacer, and repeat sequences from 10 species in the genus *Avena*. A majority of the sequences are mRNA or cDNA sequences from *A. sativa*, and these 644 sequences can be searched for SSRs. GenBank contains only 12 *Avena* genomic sequences containing introns, 2 of which are from chloroplasts. Because of the limited availability of sequence information from some crop species, such as oat, we want to develop methods to maximally exploit the available sequence data for DNA marker development.

One approach to exploiting limited sequence information is to develop PCR primers targeting specific gene regions to determine which regions provide sufficient amplification product length variation for use as DNA markers. Genomic DNA sequences (in contrast to cDNA sequences) from the public databases often indicate the positions of exons, introns, and promoter regions within the primary gene sequence. Therefore, genomic-sequence information can be used to develop PCR primers flanking the exon, intron, or promoter regions of known genes with high specificity. It may be possible to use such PCR primers as an alternative source of DNA markers that share the advantages of SSRs in variability, specificity, speed, and inexpensiveness.

DNA sequences that do not code for protein products are potentially more variable among alleles within a species because the fitness consequences of sequence variations in those regions are expected to be smaller than variations within coding sequences. Intron sequences evolve more rapidly than exon sequences in both plants (Small and Wendel 2000) and mammals (Hughes and Yeager 1997). Under the assumption that intron sequences evolve independently of function, they have been used as estimates of "genetic time" in phylogenetic studies (He and Haymer 1997; Johnson and Soltis 1994). Further, length variation among intron alleles within plant species has been reported. Chetelat et al. (1995) developed an allele-specific PCR marker for the *sucr* gene in tomato (*Lycopersicon* sp.) by designing primers to amplify several exons and introns, and demonstrated that length variations among genotypes were caused by insertion–deletion (indel) polymorphisms within an intron. Similarly, Hongtrakul et al. (1998) identified intron-length polymorphism in sunflower (*Helianthus annus* L.) caused by indels and differences in lengths of monomeric repeats, and suggested their utility as allele-specific DNA markers. Intron-length variation due to transposable-element indels in introns has also been discovered among alleles of maize genes (Bureau and Wessler 1994; Esen and Bandaranayake 1998).

The relative levels of allelic variability in lengths of introns, exons, or other gene regions have not been adequately studied in any species (Long and de Souza 1998). We expect that the phenotypic consequences of length variation are not equal among gene regions. Sequences that have large and direct effects on phenotypes are likely to maintain the least amount of variation under selection. Therefore, we expect exon sequences to have lower levels of length polymorphism, but the relative levels of variation among exons, introns, and upstream regions are not predictable because introns as well as promoter regions can affect gene expression (Bolle et al. 1996; Long and de Souza 1998). This implies that not all variation within introns and promoter regions is neutral. Introns contain the splice sites that direct their correct removal from the gene when initial transcripts are processed to mature RNAs (Lewin 1997). If these sites are changed or removed, the gene product may be altered, resulting in a nonfunctional protein (Brown et al. 1996). Laurie and Stam (1994) determined that polymorphism within an intron of the alcohol dehydrogenase gene in *Drosophila melanogaster* has an effect on the amount of protein present. Similarly, Fridman et al. (2000) suggested that length variation within the intron of an invertase gene in tomato was responsible for allelic differences in the gene's expression. Furthermore, intron size itself seems to be constrained within maximal and minimal limits by selection pressures (Carvalho and Clark 1999). The importance of promoters in gene expression is obvious, but it is not known to what extent variation within a few hundred base pairs (bp) upstream of the transcription start site will affect expression. Empirical evidence regarding the relative levels of variation among these different gene regions is needed to better understand the phenotypic importance of different gene regions and predict the feasibility of developing DNA markers by amplifying particular gene regions.

The objectives of this study were to develop PCR primer pairs targeting previously sequenced genes from *Avena* and maize in order to compare the amount of allelic amplification product polymorphism among gene regions. Exon, intron, promoter, and SSR target regions were compared in maize, while SSR and intron target regions were compared in oat.

Materials and methods

PCR primer design for maize genes

Ninety-one primer pairs were developed to target 54 GenBank

accessions representing maize nuclear genomic sequences (Table 1). Sequences were chosen at random, with the stipulation that exon and intron or noncoding 5′ regions were identified in the sequence. Intron regions were sampled more extensively because we hypothesized, before conducting the experiment, that these regions are most polymorphic. Forty-three primer pairs targeted intron sequences, 24 primer pairs targeted exon sequences, and 24 primer pairs targeted upstream 5′ flanking sequences. Exon primer pairs targeted only exons with no known intervening introns. Intron primer pairs generally targeted a small region of exon flanking either side or both sides of the intron as well as the intron itself. Two of the intron primer pairs targeted two introns and a small intervening exon. Most promoter primer pairs targeted sequences identified as promoters, TATA boxes, CAAT boxes, or specific promoter elements in GenBank (Table 1). All genes for which exons were targeted also contained promoters and (or) introns that were targeted by different primer pairs (Table 1). Two promoter primer pairs targeted a portion of the first exon of their target gene. Primers were designed with the PRIMER3 program (Rozen and Skaletskty 1997) to meet as closely as possible the optima of 20 bp primer length, 60°C annealing temperature, 20–60% GC content, and 100–200 bp amplified product size.

PCR primer design for *Avena* genes

Twenty-six PCR primer pairs were developed from 16 nuclear DNA sequences of *Avena fatua* and *A. sativa* published in GenBank (Table 2). Target sequences were either introns from a genomic DNA sequence or SSRs visually identified in either cDNA or exon sequences from genomic DNA (except that one SSR was located within an intron). SSRs contained perfect or imperfect repeats comprising at least 12 bp (Table 2). Primer pairs targeting intron sequences and a pair targeting an SSR were developed for the *GLAV1* gene; all other sequences were targeted by either intron or SSR markers, but not both (Table 2). Primers were designed with Prophet software (MarketMiner Inc., Charlottesville, Va.), based on the optimal criteria used for the maize primer pairs. Once designed, each primer was evaluated for dimer and hairpin quality with the Primer Premier program (Premier Biosof International, Palo Alto, Calif.). The primers had a Gibbs free energy of less than 0.3 kcal·mol^{-1}.

Plant material

DNA was isolated from 48 maize (*Zea mays*) inbred lines: 38-11, A12, A188, A554, A632, B14A, B37, B52, B73, B77, B84, B97, C103, CM105, CMV3, Ep1, F2, F2834T, GT112, H95, H99, HP301, Hy, I137TN, I205, I29, Ia2132, IDS28, Il677a, K55, Ky21, M37W, Mo17, N28Ht, NC258, NC296, NC298, NC304, Oh43, P39, SA24, SC213, Tx303, Tx601, Va35, W117Ht, W182B, and W64A. These lines are a subset of those surveyed by Senior et al. (1998) for SSR polymorphism, and were selected to represent as much of the genetic variation present in the original sample as possible. Maize DNA was isolated using a modified cetyltrimethlyammonium bromide (CTAB) procedure (Saghai-Maroof et al. 1984). DNA was also isolated from a single plant of each of 22 *A. sativa* cultivars (Blaze, Burton, Chaps, Clintland 64, Coker 234, Dane, Don, Horicon, Jerry, Jim, Kanota, Newdak, Ogle, Prairie, Premier, Rodeo, Skakun, Sheldon, Starter, TAM O-301, and Victoria) and 2 diploid oat accessions (*Avena strigosa* Schreb. CI1994 and *Avena wiestii* Steud. CI3815). Oat DNA was isolated using the Puregene protocol (http://www.gentra.com/Product/Protocols_Framed.html).

PCR amplification

Target SSR, intron, exon, and promoter sequences were amplified using PCR. Five microlitres (50 ng) of DNA from each line investigated was loaded into a 96-well PCR plate. Ten microlitres of polymerase solution (4.3 μL distilled water, 1.5 μL bovine serum albumin, 1.5 μL of 10× L-buffer (100 mM Tris-HCl + 15 mM $MgCl_2$ + 500 mM KCl), 0.5 μL dNTPs (100 μM each dNTP), 0.2 μL (1 U) Taq polymerase, 0.5 μL (25 ng) forward primer, and 0.5 μL (25 ng) reverse primer) was pipetted into each well of the plate and one drop of mineral oil was added to each sample. The plate was covered, placed in the thermocycler, and subjected to an initial denaturation step of 95°C for 1 min. Then the plate was cycled through three steps of 94°C denaturation for 1 min, 65°C annealing for 1 min, and 72°C extension for 2 min. After this initial cycle, the same cycle was repeated 10 more times with a 1°C decrease in the annealing temperature each time. The 10th cycle (94°C for 1 min, 55°C for 1 min, and 72°C for 2 min) was repeated 20 times, followed by a 4°C soak cycle.

Gel electrophoresis

Four percent MetaPhor (Cambrex, East Rutherford, N.J.) agarose gels containing 0.15 μg ethidium bromide/mL were used to separate the PCR amplification products. The contents of each well of the PCR plate were loaded on a gel, electrophoresed at 95 V for 4–6 h in 1× TBE buffer, and photographed under UV light to detect PCR-product polymorphisms. All maize primers that exhibited unclear banding patterns or null alleles were replicated at least twice and scored independently by two researchers. All oat PCR reactions, electrophoretic separations, and scoring were performed twice independently for each primer pair.

Data analysis

The proportion of primer pairs revealing polymorphism was computed for each class of marker. Standard errors for frequencies of polymorphism were computed as standard errors of binomial variables (Weir 1990). Frequencies of polymorphism were compared among pairs of marker groups using χ^2 tests (Snedecor and Cochran 1989).

Genotypic diversity was computed for each primer pair as

$$1 - \sum_{i}^{g} P_i^2$$

where P_i is the frequency of the ith genotype, or banding pattern (Garcia et al. 1989). When each genotype produces one band, genotypic diversity measured in this way is equivalent to gene diversity and polymorphic information content as described by Senior et al. (1998).

Mean genotypic diversity was calculated for each target class. Genotypic diversity was not normally distributed, nor was it easily transformable to normality. Therefore, nonparametric one-way analyses of variance were conducted within each species to permit statistical comparisons of mean genotypic diversity among gene regions. Kruskal–Wallis tests (Steele and Torrie 1980) were used to test for overall differences among all classes and for differences between each pair of gene-region classes. Genotypic diversity values for each maize SSR primer studied by Senior et al. (1998) were recalculated for the subset of 48 inbred lines used in this study, and these were used to compare the mean SSR genotype diversity value to those of intron, exon, and promoter markers from the current experiment.

Results

Maize markers

Four primer pairs (targeting three introns and one promoter region) failed to amplify and were excluded from further analysis. Two intron and three promoter markers were also eliminated from the dataset because they produced complex, unrepeatable banding patterns. Therefore, data were collected on 38 intron markers, 24 exon markers, and 21

Table 1. Gene sequences from maize surveyed for genotypic variation, their target regions, and genotypic diversity among 48 maize inbred lines.

GenBank accession No.	Gene	Target region*	Genotype diversity
Exon marker			
AF015268	PL transcription factor, *P1*	Exon 2 (CDS)	0.00
AF023267	Leucine-rich transmembrane protein kinase, *ltk2*	Exon 1 (5′ UTR + CDS)	0.00
AH001355	Glutathione *S*-transferase	Exon 3 (3′ UTR)	0.00
AJ005343	Ama single-subunit RNA polymerase	Exon 2 (CDS)	0.00
AJ012374	Succcinate dehydrogenase-ribosomal protein S14, *shd2-rps14*	Exon 2 (CDS)	0.00
AJ131373	High mobility group, *HMGA*	Exon 6 (CDS)	0.00
AJ131374	High mobility group, *HMGc1*	Exon 6 (CDS)	0.00
AJ132240	Eukaryotic translation initiation factor 5, *eIF 5*	Exon 2 (CDS)	0.00
AJ224847	Malate dehydrogenase	Exon 4 (CDS)	0.04
AJ238785	*Mus1* mismatch repair	Exon 9 (CDS)	0.00
D63954	Fatty acid desaturase, *FAD7*	Exon 5 (CDS)	0.00
D84409	Fatty acid desaturase, *FAD8*	Exon 5 (CDS)	0.00
E17154	Phosphoenolpyruvate carboxylase	Exon 3 (CDS)	0.00
L05934	Catalase, *Cat3*	Exon 3 (3′ UTR)	0.66
L13454	Anthocyanin regulator, *P1-Bh*,	Exon 2 (CDS)	0.54
L33244	Sucrose synthase 2, *sus2*	Exon 5 (CDS)	0.00
M24258	Amyloplast-specific transit protein	Exon 7 (CDS)	0.00
U20450	Nitrate reductase, *Zmnr2g*	Exon 2 (CDS)	0.00
U77346	Lethal leaf-spot, *lls1*	Exon 7 (CDS)	0.00
X02382	Sucrose synthase	Exon 7 (CDS)	0.00
X15544	Opaque-2, *O2*	Exon 3 (CDS)	0.00
X59546	Pyruvate decarboxylase	Exon 3 (CDS)	0.00
Y11649	Protein kinase CK2 alpha subunit	Exon 3 (CDS)	0.00
Y16041	*A1*	Exon 3 (CDS)	0.00
Intron marker			
AB001387	Ferredoxin, *Fdlll*	Intron 1	0.53
AF001012	Glossy 15	Introns 2+3	0.00
AF001012	Glossy 15	Intron 4	0.64
AF023267	Leucine-rich transmembrane protein kinase, *ltk2*	Intron 1	0.00
AH001355	Glutathione *S*-transferase	Intron A	0.48
AJ005343	Ama single-subunit RNA polymerase	Intron 4	0.00
AJ012374	Succcinate dehydrogenase-ribosomal protein S14, *shd2-rps14*	Intron 1	0.00
AJ131373	High mobility group, *HMGA*	Intron 6	0.00
AJ131374	High mobility group, *HMGc1*	Intron 4	0.00
AJ131535	Hydroxyproline-rich glycoprotein, *HRGP*	Intron 1	0.00
AJ132240	Eukaryotic translation initiation factor 5, *eIF 5*	Intron 1	0.50
AJ223471	Acidic phytase, *PHYII*	Intron 1	0.39
AJ224847	Malate dehydrogenase	Intron 1	0.43
AJ238785	*Mus1* mismatch repair	Intron 3	0.00
D00012	Triosephosphate isomerase	Introns 7+8	0.00
D45408	Translation elongation factor 1, *zmEF1A*	Intron 2	0.42
D63342	Cysteine proteinase inhibitor	Intron 2	0.49
D84409	Fatty acid desaturase, *FAD8*	Intron 5	0.00
E17154	Phosphoenolpyruvate carboxylase	Intron 4	0.00
L00371	Triosephosphate isomerase 1	Intron G	0.00
L05934	Catalase, *Cat3*	Intron 2	0.65
L29418	Sucrose synthase 1, *sus1*	Intron 1	0.38
L33244	Sucrose synthase 2, *sus2*	Intron 5	0.60
L33244	Sucrose synthase 2, *sus2*	Intron 10	0.00
M73235	*O*-Methyltransferase, OMT	Intron 1	0.50
U09989	D3L H(+)-transporting ATPase, *Mha1*	Intron 3	0.76
U15964	Xyloglucan endo-trans-glycosylase homolog	Intron 1	0.28
U20450	Nitrate reductase, *Zmnr2g*	Intron 2	0.12

Table 1 (*concluded*).

GenBank accession No.	Gene	Target region*	Genotype diversity
U28017	Globulin, *Glb1*	Intron 2	0.45
U77346	Lethal leaf-spot, *lls1*	Intron 5	0.00
U93178	Helix–loop–helix-type transcription factor, *R*	Intron 1	0.08
X02382	Sucrose synthase	Intron 5	0.00
X15544	Opaque-2, *O2*	Intron 4	0.39
X59546	Pyruvate decarboxylase	Intron 2	0.64
Y11403	Cytochrome P450, *CYP71C3*	Intron 2	0.50
Y11649	Protein kinase CK2 alpha subunit	Intron 1	0.28
Y16041	*A1*	Intron 1	0.63
Z33612	Phosphoglycerate mutase	Intron 3	0.00
Promoter marker			
AF015268	PL transcription factor, *P1*	Upstream (promoter)	0.12
AJ131535	Hydroxyproline-rich glycoprotein, *HRGP*	Upstream (including TATA)	0.00
AJ132240	Eukaryotic translation initiation factor 5, *eIF 5*	Upstream (promoter)	0.00
AJ224847	Malate dehydrogenase	Upstream (including TATA)	0.54
AJ238786	Mus2 mismatch repair	Upstream	0.00
D63342	Cysteine proteinase inhibitor	Upstream (including CAAT)	0.28
L29418	Sucrose synthase 1, *sus1*	Upstream (promoter)	0.12
L32898	Sucrose synthase 1, *sus1*	Upstream (promoter)	0.00
L40803	Glyceraldehede-3-phosphate dehydrogenase	Upstream (promoter)	0.08
X05068	*A1*	Upstream (including CAAT, TATA, and enhancer)	0.12
X15596	Glyceraldehyde-3-phosphate dehydrogenase, *GADPH*	Upstream (including anaerobic regulatory element and pyrimidine box)	0.12
X15642	Phosphoenolpyuvate carboxylase	Upstream	0.00
X56737	Auxin-binding protein, *Aux311*	Upstream (including TATA)	0.49
X60135	Catalase-1, *cat-1*	Upstream (promoter)	0.42
X60205	Chalcone synthase, *c2*	Upstream (promoter)	0.41
X78988	Proteinase inhibitor, *MPI*	Upstream (promoter + exon)	0.00
X81828	Cytochrome P-450, *CYP71C1*	Upstream	0.00
Y00322	Ribulose bisphosphate carboxylase small subunit	Upstream (including TATA)	0.19
Y11649	Protein kinase CK2 alpha subunit	Upstream (including TATA + exon)	0.25
Z26824	Ferredoxin-NADP reductase binding protein	Upstream (including TATA)	0.28
Z54358	Catalase, *Cat2*	Upstream (promoter)	0.49

*CDS, coding sequence (translated region); UTR, untranslated region.

promoter markers (Table 1). Three (8%) intron markers and one (4%) exon marker produced more than two repeatable bands in some genotypes, indicating that these primer pairs amplified two or more duplicate loci. Six (16%) intron markers and five (24%) promoter markers produced presence or absence of amplification product polymorphisms. These proportions were significantly greater than that for exon markers, none of which produced presence or absence polymorphisms.

Polymorphisms were detected significantly more frequently for promoter and intron markers (67 and 58%, respectively) than for exon markers (13%; Table 3). Taking all primer pairs into account, mean genotype diversities of promoter and intron markers were similar to each other and significantly greater than diversity of exon markers (Table 3). Limiting the comparisons to polymorphic primer pairs only (Table 4) revealed that polymorphic intron markers had greater mean genotypic diversity than polymorphic promoter markers (Table 3). Because there were only three polymorphic exon primers, reliable comparison of genotype diversity of polymorphic exon markers with other polymorphic markers was hindered.

Genotype diversity of 70 SSR markers among the 48 maize inbred lines assayed in this experiment was calculated from the data of Senior et al. (1998), revealing that genetic diversity of SSRs was significantly greater than that of either intron or promoter markers (Table 3). We could not directly compare the proportion of polymorphic SSR markers to intron, promoter, or exon markers in maize, however, because of the method used to select the maize SSR markers. Chin et al. (1996) performed an initial screening for polymorphism using nine inbred lines only, including four proprietary inbred lines. They reported that 69 of 200 (35%) maize SSRs were polymorphic on their set of nine inbred lines. The more stringent selection of SSRs for polymorphism on a smaller set of lines may have resulted in maintaining only the most polymorphic SSRs to use in assays of genetic diversity. Therefore, the mean genotypic diversity

Table 2. Gene sequences from *Avena* surveyed for genotypic variation, their target regions, and genotypic diversity among 22 hexaploid oat cultivars and two diploid oat accessions.

GenBank accession No.	Gene	Target region*	Genotypic diversity
SSR marker			
AF033096	Non-phototrophic hypocotyl, *NPH1-1*	SSR, $(AGG)_{3+1}$-$(GGC)_5$-$(TCC)_3$.$(GCC)_2$ (5′ UTR)	0.00
AF033096	Non-phototrophic hypocotyl, *NPH1-1*	SSR, $(ATGTA)_3$ (3′ UTR)	0.46
AF033097	Non-phototrophic hypocotyl, *NPH1-2*	SSR, $(TGTA\text{-})_3$ (3′ UTR)	0.45
AF033097	Non-phototrophic hypocotyl, *NPH1-2*	SSR, $(GCT)_{3+1+1}$-$(GGC)_5$ (5′ UTR)	0.00
AJ133638	Myba	SSR, $(GAGA)_4(TCC\text{-})_4$ (5′ UTR)	0.00
L39777	Thaumatin-like pathogenesis-related protein, *Rast1-4*	SSR, $(AAAT)_3$ (3′ UTR)	0.57
M18822	Type 3 phytochrome gene, *Phy3*	SSR, $(TA)_6$ (intron B)	0.00
M18822	Type 3 phytochrome gene, *Phy3*	SSR, $(ATT\text{-})_3$ (CDS)	0.13
M83381	Avenin	SSR, $(GCA)_4(CAA)_5$-$(AGC\text{-})_2$ (CDS)	0.44
X03244	Phytochrome *AP5* fragment	SSR, $(CAC\text{-})_4$ (CDS)	0.14
X74740	11S globulin, *GLAV 1*	SSR, $(CTAACAG)_3$ (5′ UTR)	0.36
Z48431	DNA binding protein (*A. fatua*)	SSR, $(CCA\text{-})_2(GCC)_2(CAG)_4$-$(GTC)_4$ (CDS)	0.13
Z48429	DNA binding protein (*A. fatua*)	SSR, $(GGC)_5$-$(GCG)_5$ (CDS)	0.50
Z83832	UDP-glucose:sterol glucosyltransferase	SSR, $(TCG\text{-})_3(GTT\text{-})_6$ (3′ UTR)	0.00
Intron marker			
AJ010728	Alpha-amylase, *Amy2A* gene (*A. fatua*)	Intron 3	0.70
AJ010729	Alpha-amylase, *Amy2D* gene (*A. fatua*)	Intron 1	0.46
AJ010729	Alpha-amylase, *Amy2D* gene (*A. fatua*)	Intron 2	0.34
X17637	Seed-storage protein, *SSP12*	Intron 1	0.00
X17637	Seed-storage protein, *SSP12*	Intron 2	0.46
X17637	Seed-storage protein, *SSP12*	Intron 3	0.13
X68648	12S seed globulin pseudogene	Intron 1	0.00
X68648	12S seed globulin pseudogene	Intron 2	0.25
X74740	11S globulin, *GLAV 1*	Intron 1	0.00
X74740	11S globulin, *GLAV 1*	Intron 4	0.00
X74741	11S globulin, *GLAV 3*	Intron 1	0.00
X74741	11S globulin, *GLAV 3*	Intron 2	0.79

Note: Genes are from *A. sativa* unless noted otherwise as being from *A. fatua*.

*CDS, coding sequence (translated region); UTR, untranslated region. Subscripts following repeat sequences refer to the number of repeats in the reference sequence; a subscript $n + 1$ refers to a sequence that is repeated n times, interrupted by fewer than five bases and repeated again; a hyphen within the parentheses denotes a variable base of an imperfect repeat.

measure for SSRs reported here may be biased upward relative to the other markers because of preselection.

Avena markers

The proportions of markers that were polymorphic in the sample of *Avena* lines tested were approximately equal for SSR (64%) and intron (58%) markers (Table 3). No statistically significant differences were observed between *Avena* intron and SSR markers for genetic diversity, regardless of whether all primer pairs or only polymorphic primer pairs were used in the calculations (Table 3). The genotypic diversity of polymorphic *Avena* primer pairs (Table 5) was 0.45 for intron markers and 0.35 for SSRs. The most polymorphic primer pairs were the SSRs *Rast1-4* and *NPH1-1-2* and the intron markers *AMY2A*, *GLAV3* intron 2, and *SSP12* intron 2.

If only the hexaploid cultivars are considered, the percentage of polymorphism drops to 21% for the SSRs and 50% for the intron markers. Similarly, Li et al. (2000) observed that whereas 62% of their SSRs were polymorphic when tested on a set of 12 *Avena* species and 20 oat cultivars, only 36% were polymorphic within the set of cultivars. Our results cannot be directly compared with those of Li et al. (2000) because of the different samples of genotypes assayed and because most (89%) of the SSR markers developed here revealed presence or absence polymorphisms rather than only amplification product length variation, as do typical SSRs. Furthermore, we may have underestimated the proportion of polymorphic markers in this study by using MetaPhor agarose gels, whereas Li et al. (2000) used longer polyacrylamide gels to improve discrimination of similarly sized alleles.

Seven of nine polymorphic *Avena* SSRs produced one or fewer bands per genotype and two *Avena* SSRs produced two or fewer bands per genotype. Polymorphic intron markers produced from zero to three bands per genotype. Without direct genetic analysis of the inheritance of these bands, it was not clear whether multiple bands per genotype were due to residual heterozygosity or to duplicated loci amplified by

Table 3. Proportion of polymorphic markers and mean genotypic diversity values (including or excluding non-polymorphic markers) for different marker types in oat and maize.

Target sequence type	No. of primer pairs	Proportion of polymorphic primer pairs* (%)	Genotype diversity for all primer pairs	Genotype diversity for polymorphic primer pairs
Avena genes				
Intron	12	58 ± 14	0.26	0.45
SSR	14	64 ± 13	0.23	0.35
Significance (Kruskal–Wallis test)	ns	ns		
Maize genes				
Exon	24	13 ± 6*b*	0.05*b*	0.42*abc*
Intron	38	58 ± 8*a*	0.27*a*	0.46*b*
Promoter	21	67 ± 10*a*	0.19*a*	0.28*c*
SSR	70	na	na	0.60*a*
Significance (Kruskal–Wallis test)	0.001	0.0001		

Note: Values followed by the same letter are not significantly different at the 0.05 probability level, based on the Kruskal–Wallis test; na, not applicable; ns, not significant.

*Mean ± SE.

the same primer pair. The allohexaploid nature of the *A. sativa* genome and the high rate of self-pollination in oat makes the existence of multiple loci likely.

Seed-storage protein (SSP) 12 intron 3 marker (from accession X17637) was polymorphic between CI3815 and CI1994, the parents of a diploid oat mapping population. Therefore, the *A. strigosa* × *A. wiestii* recombinant inbred mapping population was assayed with this marker. The marker segregated as a single locus, *SSP12-3*, and was mapped by Kremer et al. (2001) in this population. The *Rast1-4* (L39777) SSR marker and *Glav3* (X7471) intron 2, *Amy2A* (AJ10728) intron 3, and *Amy2D* (AJ10729) intron 1 markers were polymorphic between hexaploid oat cultivars Ogle and TAM O-301 and were mapped in the recombinant inbred mapping population developed from the cross of these cultivars. The *Glav3* intron 2 marker identified two linked loci (*Glav3.1* and *Glav3.2*) 9 cM apart (Portyanko et al. 2001). *Rast1-4*, *Amy2A*, and *Amy2D* mapped as single loci (Portyanko et al. 2001).

Discussion

Our results indicate that intron markers exhibit substantial within-species variation and can be a useful source of DNA markers for oat and maize breeding and genetics. Promoter markers are also frequently polymorphic among maize inbred lines; however, the genotypic diversity revealed by those polymorphic markers is less than that revealed by polymorphic intron markers. Intron and promoter markers share desirable properties with SSRs, including PCR-based assays, locus specificity, and relatively high polymorphism, except that they detected less genetic diversity than the selected set of SSRs used for comparison in maize. Intron and promoter markers can be more difficult than SSRs to develop, however, because they can be developed only for genes with well-annotated genomic sequences, whereas SSRs can be detected in both genomic and cDNA sequences. Fewer genomic than cDNA sequences are found in the public databases. For example, as of July 2001, GenBank contained 767 DNA sequences from *Avena*, of which only 10 were genomic nuclear DNAs containing introns. Fewer than 1% of maize and *Arabidopsis* DNA sequences in GenBank contained introns at that time. The number of genomic-sequence resources for major crop species is expected to increase dramatically in the future, which will facilitate development of intron and promoter markers. cDNA sequences will likely be added at an even faster rate, however, because of expressed sequence tag projects under way for many plant species (Davis et al. 2000; Lonsdale et al. 1999).

Our comparisons involving oat SSR markers are confounded by the finding that only one of the polymorphic oat SSR markers (X03244) demonstrated solely length polymorphism, while all others revealed presence or absence polymorphisms (Table 5). This indicates that the sequence variation underlying most of the polymorphisms was most likely in the sequences homologous to the primers, which flank the repeats (Gupta and Varshney 2000). This is in contrast to typical SSR polymorphisms, which represent varying numbers of the targeted repeat sequence, specified as simple sequence length polymorphisms by Cho et al. (2000). Searches of cDNA sequences for short repeat sequences, like the one performed in this study, may be less likely to result in identifying SSRs with high levels of repeat-length polymorphism than the approach taken by Li et al. (2000) of identifying longer SSRs in hybridization-enriched libraries. SSR markers developed from genomic libraries were found to be substantially more polymorphic than those from expressed sequence tag libraries in rice (Cho et al. 2000) and durum wheat (Eujayl et al. 2001).

Many of the cDNA sequences from which the oat SSR primers were derived represent seed-storage or DNA-binding proteins, which tend to have short stretches of repeated amino acids, and these amino acid repeats are not representative of typical SSR repeats. The number of such repeats may be conserved, owing to their effects on protein function, and therefore less likely to vary than repeat sequences in noncoding regions. This may explain why, among hexaploid oat cultivars, the repeat markers used in this study were significantly less polymorphic than intron markers. The repeat markers developed here seem to have little advantage over those developed by Li et al. (2000) except that they were easy to create and they provide allelic tags for known genes.

Table 4. GenBank accession Nos., DNA sequences, total numbers of unique amplification product sizes, range of bands per genotype, and numbers of unique genotypes of polymorphic primer pairs from maize.

GenBank accession No.	Primer pair DNA sequences, 5′ → 3′*	No. of unique amplification products	No. (or range) of bands per genotype†	No. of unique banding patterns
Exon marker				
AJ224847	GAGAAATGAGAGGCTATTTTACAAGC TGTATAAACCCTGCGGTCGT	2	1–2	2
L05934	GTCGCTGGGGATGAAGATT GGTACAAAGGCGGACACG	3	1	3
L13454	CTGCGGTGGCTGAACTACCT CTGTTGCCGAGGAGCTTGT	3	1–3	4
Intron marker				
AB001387	CACCAGGACCCTTGTTCG ACCATCCATCCACTCGACTT	2	0–1	3
AF001012	CAATTGCGAATCGAACGAC CACGGCAGGAAGACAAGAAT	4	1–4	6
AH001355	CCAGTTGATTGTCAATTGTTTTG GGCGCATACAAGAACCAGAT	1	0–1	2
AJ132240	ATAGCCCGTTCATCGATCC TTCCTGCGGATACTCAAACC	2	1	2
AJ223471	TCGCTTTGATAACATGATCTGC CCATATATACTAAGACAAATCGTTGCT	2	1	2
AJ224847	GTATGCGGGTTTCATTCTGC TGAAATCCCGGTAATGACAAA	2	1	2
D45408	AGCATGCTCTCCTTGCTTTC CAATCACAGCACAAATGCAG	3	1–3	3
D63342	CGAAGCTAAGGTCTGGGAGA CTCTCCCAACGACTGTTGAA	2	1	2
L05934	GCAGTGCCACACGTGTTTT GCATATGCAAAGCAATCGTC	2	0–2	4
L29418	GCTCTAGCCCTCTCCTCCTC CATGACCTCCCCTACACCAT	3	1–3	4
L33244	TGAGCCATTATTTTATCTTTTTCTTT	2	1–2	3
Intron 5	CTGTAGGGAACCATAGAATTTCA			
M73235	AGTAGTAGCCGCATCGCATC CTAGCCGCAGCCAGTAGAGT	2	1	2
U09989	CCTCAAATTTTTGGGCTTC TCAACGGAAACACACATCA	4	1–2	6
U15964	ACATGCAGCTCAAGCTCGTC CGCGCGAGAGAGAGAGAC	2	0–1	3
U20450	GATGTGGAGTGGCAACTGG TGCAACAAGTAGGAACGATGA	1	0–1	2
U28017	TTATTAAGCTCTAAACGAGACATCTG CTGCCAAAGGAACGAAGG	2	1	2
U93178	CAGTTGATATTTGGTTGGCAGT AAAACCTGGCGACATCATTG	3	1–2	3
X15544	CATCGATCGTCGTTCCATTT GGCAAAACTGAACCCAAAAA	2	1	2
X59546	AATACGGCACATCCATCCA GATCGTCCGATCAGCAGTGT	3	1	3
Y11403	GGCGACAAGGAGAAGGAGTC GGCCACATGGAAAGGAAGT	3	1	3
Y11649	GCTCTCCTCTCGCGTCCT GCAATCAATGAGCCTAATGGT	2	1	2
Y16041	TACGTACGCAGCCACTTGTC GGGGTGGACATAAATAAAACG	4	0–1	5

Table 4 (*concluded*).

GenBank accession No.	Primer pair DNA sequences, 5′ → 3′*	No. of unique amplification products	No. (or range) of bands per genotype†	No. of unique banding patterns
Promoter marker				
AF015268	TTTACCAATGTTGACATTCATCG GAGGGTGGCAACTGTCAGATA	2	0–2	2
AJ224847	TCGCTAGTAGAAGACAGAATTAAAAA ACACGTTGCAGGAGAGGTG	2	0–1	3
D63342	TCTCTTATCATTTTCTCACCTGAA CGGAAATTTATAGCCGGAGA	2	1	2
L29418	CAGCCGTGTGTCTCTTCAAA CGTCACCTTCTGACCTTGGTA	3	1–2	3
L40803	CACCCTCAATGTCCACAGAA TTTCGAGAAGCTTCCCATTG	2	1	2
X05068	GGCTCTAGCTAGGCGTGTCA AAGGTACGCAGGCAGAAAAA	2	1	2
X15596	CCGCACCTCATTCCACAA AAAGGAGAAAACGAAATCCAACTAC	3	1–2	3
X56737	AATCATTCCGGGTGCTCA CAGATAGCCTGTGGGAGCAG	1	0–1	2
X60135	GGCTAAAGAGCGAGCCAAAT GTCGACTGGCAAGGGAGATA	2	1–2	3
X60205	AACGGTTTTCCGGTCTTTTT GAGTCAGCGTCTCCCTCATC	2	1	2
Y00322	TATATATGCCGTCGGTGTGG TGCCTGGCTGCCTAGTATGT	2	1	2
Y11649	GGATTTAGACGACGGATTACG GGCAGGCCATGTAACTTTTG	2	1	2
Z26824	GAGTGATCCAAACGCCTCAT ACTGCCAATGGAGAGCTGAC	1	0–1	2
Z54358	TACAGCAGAACCCACCCAGT AGGCTGCGACACGGATAG	2	0–1	3

*In each case the forward primer sequence is listed first followed by the reverse primer sequence.
†Excluding monomorphic bands.

Most polymorphic markers targeting oat introns also exhibited presence or absence variation, but this proportion (57%) was still significantly lower than that of the oat repeat markers (89%). In contrast, in maize, none of the exon primers, 27% of the polymorphic intron markers, and 36% of the polymorphic promoter markers exhibited presence or absence polymorphisms. The difference between maize and oat intron markers may be due to the bias introduced by the high proportion of seed storage and DNA binding protein genes in our limited sample of genes targeted in oat.

Within exons the position of the amplification target region may influence the level of polymorphism. cDNA sequences can contain 5′ and 3′ untranslated regions (UTRs) in addition to translated regions. Perhaps UTRs possess a greater potential for sequence-length variability than translated regions. Scott et al. (2000) reported that the relative level of variability among 5′ UTR-, 3′ UTR-, and coding-sequence-derived SSRs depended on the level of taxonomic divergence among the small sample of grape genotypes in their study. Among the exon markers in maize, only three included any UTRs in the target sequence. One of these (from GenBank accession L05934) was polymorphic, whereas only 3 of 21 exon markers that targeted translated regions were polymorphic. Among 14 oat repeat sequences targeted, 13 were within exons, and of these, 5 were in a translated coding sequence, 4 in a 5′ UTR, and 4 in a 3′ UTR (Table 1). Surprisingly, all of the SSRs within coding sequences were polymorphic and three of four SSRs within 3′ UTRs were polymorphic; only one of four SSRs within a 5′ UTR was polymorphic. This observed distribution of polymorphism deviates significantly ($P < 0.025$) from a random distribution of polymorphisms across the three classes of exon regions, but a larger sample of genes is needed to verify this result. It is unclear why 5′ UTRs would be more conserved than coding sequences or 3′ UTRs; perhaps 5′ UTRs play an important role in gene expression. Alternatively, it may be an artefact of sampling that the repeat-sequence markers within 5′ UTRs were more conserved. For example, the frequency of trinucleotide repeats versus dinucleotide and tetranucleotide repeats may influence the level of polymorphism of repeat sequences within coding regions. The addition or deletion of one trinucleotide repeat will not result in a frame-shift mutation, whereas such polymorphisms within dinucleotide and tetranucleotide repeat sequences will result in frame-shift mutations. This appears not to have been the case in this sample, however, because (*i*) most of the

Table 5. GenBank accession Nos., DNA sequences, total numbers of unique amplification product sizes, range of bands per genotype, and numbers of unique genotypes of polymorphic primer pairs from *Avena*.

GenBank accession No.	Primer pair DNA sequences, 5′ → 3′*	No. of unique amplification products	No. (or range) of bands per genotype†	No. of unique banding patterns
SSR marker				
AF033096	TGCATGTTTTTGTTTGTTGTTG	2	0–1	3
$(ATGTA)_3$	CACGATCCAAATACACGCAG			
AF033097	GCTGAGTTTATGTTGGTGCG	1	0–1	2
$(TGTA\_)_3$	AATTTTCTCCAACCCCATCC			
L39777	CTTCTGCCCATGAAACCCTA	2	0–2	3
	ACTCAGCACATGCACCCTC			
M18822	AAGAGCTTGCCTTGGAGTGA	2	0–1	2
$(ATT\_)_3$	CAGCAAGCCCATCAAGTTTT			
M83381	ATCTGTCAGGTGACGAGGCA	2	0–2	3
	CCTTGCATCTGAGGTTGGTT			
X03244	ATGCTTACAACGGTCAGCCA	2	1	2
	TGAACAATGGCATAGAAAGGC			
X74740	GCATGATCGAACGAGTTGTG	2	0–1	3
	TTTGAGAGGTTGGTGGTAGC			
Z48431	CAGCAACAACAACAACCACC	1	0–1	2
	CACTGGTAGCCGTCCTTGAC			
Z48429	ATTCGTTCGGAGCGATGTC	1	0–1	2
	GCCGTCTTTCCTCCATTTCT			
Intron marker				
AJ010728	GCTACGCCTACATCCTCACC	3	1–3	6
	CGCGATCTCTTCCTTGAATC			
AJ010729	TGATGGGGATCAAGAAGAGC	1	0–1	2
Intron 1	ATATCGTCGACCTTCCCCAT			
AJ010729	GGCTACGCCTACATCCTCAC	2	1–2	2
Intron 2	CTCCTTGAATCCCCAGTTGA			
X17637	TTCGACGTAAACAACAACGC	2	0–1	2
Intron 2	TGTTACCAGCCAACAAGAACTC			
X17637	CCTGGTGCAAATGAGTGCTA	2	1–2	2
Intron 3	TGCATGACACTGTGAGCATT			
X68648	ACCAAAGAGTTCACCGCATC	1	0–1	2
Intron 2	TGTTACCAGCCAACAAGAACTC			
X74741	GCATTGTGCATTGGTGCTAC	3	0–2	6
Intron 2	TGTTACCAGCCAACAAGAACTC			

*In each case the forward primer sequence is listed first followed by the reverse primer sequence.
†Excluding monomorphic bands.

polymorphisms observed were presence or absence variations rather than repeat-number variations; (*ii*) there was no relationship between repeat length and polymorphism in this study; and (*iii*) we observed one tetranucleotide repeat (X03244) within a coding sequence that exhibited length polymorphism (Tables 2 and 5).

The greater level of sequence polymorphism within noncoding gene regions relative to coding regions reported here can be compared with nucleotide substitution rate data collected from different gene regions in mammals (Li 1997, pp. 182–184). The average nucleotide-substitution rate in introns is similar to that in pseudogenes and fourfold degenerate sites, and much greater than the substitution rates in non-degenerate coding triplets (Li 1997), which is similar to our result for sequence-length polymorphism in maize and oat. Mammalian 5′ flanking, 5′ UTR, and 3′ UTR substitution rates are intermediate to intron and non-degenerate coding sequences, in contrast to our findings in maize that promoter regions are highly polymorphic for sequence length and that less polymorphism was observed in oat SSRs in 5′ UTRs than in 3′ UTRs. Direct comparison of large segments of DNA sequence from two *Arabidopsis* genotypes revealed that single nucleotide polymorphisms occurred at a frequency 1.4 times greater and indel polymorphisms at a frequency 3 times greater in introns than in exons (The Arabidopsis Genome Initiative 2001). Promoter and intron gene regions are not necessarily selectively neutral sequences, but they have significantly greater length polymorphism within species than do exon regions. In contrast, our limited data on polymorphism within 5′ UTRs of exons suggest that these regions have a greater effect on fitness (perhaps because of mRNA stabilization and transport). Thus, variation in some, but not all, noncoding regions has sufficiently small fitness consequences, so plant populations can

maintain high levels of sequence variation within these regions. The practical consequence of this is that for crop species with limited polymorphic PCR marker availability (like oat), public sequence databases can be easily and successfully mined to develop polymorphic markers.

References

Bolle, C., Herrmann, R.G., and Oelmuller, R. 1996. Intron sequences are involved in the plastid- and light-dependent expression of the spinach PsaD gene. Plant J. **10**: 919–924.

Brown, J.W.S., Smith, P., and Simpson, C.G. 1996. *Arabidopsis* consensus intron sequences. Plant Mol. Biol. **32**: 531–535.

Bureau, T.E., and Wessler, S.R. 1994. Stowaway: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. Plant Cell, **6**: 907–916.

Carvalho, A.B., and Clark, A.G. 1999. Intron size and natural selection. Nature (London), **401**: 344.

Chen, X., Temnykh, S., Xu, Y., Cho, Y.G., and McCouch, S.R. 1997. Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). Theor. Appl. Genet. **95**: 553–567.

Chetelat, R.T., DeVerna, J.W., and Bennett, A.B. 1995. Introgression into tomato (*Lycopersicon esculentum*) of the *L. chmielewskii* sucrose accumulator gene (*sucr*) controlling fruit sugar composition. Theor. Appl. Genet. **91**: 327–333.

Chin, E.C.L., Senior, M.L., Shu, H., and Smith, J.S.C. 1996. Maize simple sequence repetitive DNA sequences: abundance and allele variation. Genome, **39**: 866–873.

Cho, Y.G., Ishii, T., Temnykh, S., Chen, X., Lipovich, L., McCouch, S.R., Park, W.D., Ayres, N., and Cartinhour, S. 2000. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet. **100**: 713–722.

Davis, G.L., McMullen, M.D., Baysdorfer, C., Musket, T., Grant, D., Staebell, M., Xu, G., Polacco, M., Koster, L., Melia-Hancock, S., Houchins, K., Chao, S., and Coe, E.H., Jr. 2000. A maize map standard with sequenced core markers, grass genome reference points and 932 expressed sequence tagged sites (ESTs) in a 1736-locus map. Genetics, **152**: 1137–1172.

Esen, A., and Bandaranayake, H. 1998. Insertional polymorphism in introns 4 and 10 of the maize β-glucosidase gene *glu1*. Genome, **41**: 597–604.

Eujayl, I., Sorrells, M.E., Baum, M., Wolters, P., and Powell, W. 2001. Assessment of genotypic variation among cultivated durum wheat based on EST-SSRs and genomic SSRs. Theor. Appl. Genet. In press.

Fridman, E., Pleban, T., and Zamir, D. 2000. A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. Proc. Natl. Acad. Sci. U.S.A. **97**: 4718–4723.

Garcia, P., Vences, F.J., Perez de la Vega, M., and Allard, R.W. 1989. Allelic and genotypic composition of ancestral Spanish and colonial Californian gene pools of *Avena barbata*: evolutionary implications. Genetics, **122**: 687–694.

Gupta, P.K., and Varshney, R.K. 2000. The development and use of microsatellite markers for genetics analysis and plant breeding with emphasis on bread wheat. Euphytica, **113**: 163–185.

He, M., and Haymer, D.S. 1997. Polymorphic intron sequences detected within and between populations of the oriental fruit fly (Diptera: Tephritidae). Ann. Entomol. Soc. Am. **90**: 825–831.

Hongtrakul, V., Slabaugh, M.B., and Knapp, S.J. 1998. DFLP, SSCP, and SSR markers for delta 9-stearoyl-acyl carrier protein desaturases strongly expressed in developing seeds of sunflower: intron lengths are polymorphic among elite inbred lines. Mol. Breed. **4**: 195–203.

Hughes, A.L., and Yeager, M. 1997. Comparative evolutionary rates of introns and exons in murine rodents. J. Mol. Evol. **45**: 125–130.

Johnson, L.A., and Soltis, D.E. 1994. *matK* DNA sequences and phylogenetic reconstruction in Saxifragaceae s.str. Syst. Bot. **19**: 143–156.

Kremer, C.A., Lee, M., and Holland, J.B. 2001. A restriction fragment length polymorphism based linkage map of a diploid *Avena* recombinant inbred line population. Genome, **44**: 192–204.

Laurie, C.C., and Stam, L.F. 1994. The effect of an intronic polymorphism on alcohol dehydrogenase expression in *Drosophila melanogaster*. Genetics, **138**: 379–385.

Lewin, B. 1997. Genes VI. Oxford University Press, Oxford.

Li, C.D., Rossnagel, B.G., and Scoles, G.J. 2000. The development of oat microsatellite markers and their use in identifying relationships among *Avena* species and oat cultivars. Theor. Appl. Genet. **101**: 1259–1268.

Li, W.-H. 1997. Molecular evolution. Sinauer Associates, Sunderland, Mass.

Long, M., and de Souza, S.J. 1998. Intron–exon structures: from molecular to population biology. *In* Advances in genome biology, genes and genomes. *Edited by* R.S. Verma. JAI Press, London. pp. 143–178.

Lonsdale, D., Arnold, B., and Arnold, B. 1999. Mendel-ESTS: database of plant ESTs in dbEST annotated with gene family numbers and gene family names. Plant Mol. Biol. Rep. **17**: 239–247.

Portyanko, V.A., Hoffman, D.H., Lee, M., and Holland, J.B. 2001. A linkage map of hexaploid oat based on grass anchor DNA clones and its relationship to other oat maps. Genome, **44**: 249–265.

Saghai-Maroof, M.A., Soliman, K.M., Jorgensen, R.A., and Allard, R.W. 1984. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc. Natl. Acad. Sci. U.S.A. **81**: 8014–8018.

Scott, K.D., Eggler, P., Seaton, G., Rossetto, M., Ablett, E.M., Lee, L.S., and Henry, R.J. 2000. Analysis of SSRs derived from grape ESTs. Theor. Appl. Genet. **100**: 723–726.

Senior, M.L., and Heun, M. 1993. Mapping maize microsatellites and polymerase chain reaction confirmation of the targeted repeats using a CT primer. Genome, **36**: 884–889.

Senior, M.L., Murphy, J.P., Goodman, M.M., and Stuber, C.W. 1998. Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. Crop Sci. **38**: 1088–1098.

Small, R.L., and Wendel, J.F. 2000. Copy number lability and evolutionary dynamics of the *Adh* gene family in diploid and tetraploid cotton (*Gossypium*). Genetics, **155**: 1913–1926.

Smith, J.S.C., Chin, E.C.L., Shu, H., Smith, O.S., Wall, S.J., Senior, M.L., Mitchell, S.E., Kresovich, S., and Ziegle, J. 1997. An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPs and pedigree. Theor. Appl. Genet. **95**: 163–173.

Snedecor, G.W., and Cochran, W.G. 1989. Statistical methods. 8th ed. Iowa State University Press, Ames.

Steele, R.G.D., and Torrie, J.H. 1980. Principles and procedures of statistics. 2nd ed. McGraw-Hill, New York.

Taramino, G., and Tingey, S. 1996. Simple sequence repeats for germplasm analysis and mapping in maize. Genome, **39**: 277–287.

The Arabidopsis Genome Initiative. 2001. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature (London), **408**: 196–815.

Wang, Z., Weber, J.L., and Tanksley, S.D. 1994. Survey of short tandem repeats. Theor. Appl. Genet. **88**: 1–6.

Weir, B.S. 1990. Genetic data analysis. Sinauer Associates, Sunderland, Mass.

Westman, A.L., and Kresovich, S. 1998. The potential for cross-taxa simple-sequence repeat (SSR) amplification between *Arabidopsis thaliana* L., and crop brassicas. Theor. Appl. Genet. **96**: 272–281.