Theor Appl Genet (2002) 105:532–543
DOI 10.1007/s00122-002-0897-1

R. Mogg · J. Batley · S. Hanley · D. Edwards
H. O'Sullivan · K.J. Edwards

Characterization of the flanking regions of *Zea mays* microsatellites reveals a large number of useful sequence polymorphisms

Received: 4 September 2001 / Accepted: 10 December 2001 / Published online: 23 May 2002

Abstract Sequence characterization of the flanking regions of 52 sequence-tagged microsatellite loci and two gene fragments from 11 *Zea mays* inbred lines identified a total of 324 sequence polymorphisms. The sequence polymorphisms consisted of both single-nucleotide polymorphisms and insertions/deletions in a ratio of approximately two to one. The level of sequence variation within the flanking regions of microsatellites linked to expressed sequence tags was lower than microsatellites that were unlinked to expressed sequence tags. However, both types of microsatellites generated a similar number of sequence-based alleles across the 11 genotypes surveyed. In two out of 20 microsatellites examined in detail, evidence was found for size-based allele homoplasy. Conversion of the observed sequence polymorphisms into allele-specific oligonucleotides followed by covalent binding to glass slides allowed the sequence polymorphisms to be used in a simple hybridization-based genotyping procedure. This procedure enabled us to discriminate between different inbred lines and allowed variations within a single inbred to be identified. The sequence information presented in this report could be used as a starting point for other programmes in the further development of a non-gel based, multi-locus, multi-allele screen for large-scale maize genotyping.

Keywords Microsatellites · SSRs · *Zea mays* · Single nucleotide polymorphism · Indels

Introduction

Marker-assisted breeding and genome mapping both rely upon the availability of polymorphic genetic markers. Such markers include restriction fragment length polymorphisms (RFLPs; Helentjaris et al. 1986), amplified fragment length polymorphisms (AFLPs; Vos et al. 1995) and sequence-tagged-microsatellites or simple sequence repeats (SSRs; Weber and May 1989). Recently, SSR markers have become the marker of choice for molecular plant breeders (Gupta et al. 1996). SSR markers are co-dominant, highly polymorphic and multi-allelic. Unfortunately, the methods for detecting both SSRs and the other types of polymorphic markers, rely upon the electrophoretic separation of DNA in agarose or polyacrylamide gels. For example, at an individual SSR locus, the variation in allele fragment size arising from differences in the number of repeat units can be detected by a combination of the polymerase chain reaction (PCR) and denaturing polyacrylamide gel electrophoresis (Sambrook et al. 1989). Developments in fluorescent DNA fragment analysis have made it possible to both analyze many SSR loci simultaneously and automatically capture the resulting data. However, despite the advent of these semi-automated systems and refinements such as capillary gel electrophoresis (Gonen et al. 1999), gel-based technology is still labour-intensive and time-consuming for the large-scale genotyping required in experimental genome analysis, marker-assisted breeding programmes and linkage disequilibrium studies (Brookes 1999).

The requirements for a high-throughput genotyping system might include increased scope for automation and a simple binary scoring system that can be reliably read by machine, with no human intervention. The differential hybridization between probe DNA and allele-specific oligonucleotides (ASOs) which underpin so-called 'DNA genotyping chips' could provide the basis for such a system. ASO technology is based upon the principle that when hybridized under appropriate conditions, synthetic DNA oligonucleotide probes (15–25 bases) will anneal to their complementary PCR-generated target sequences only if they are perfectly matched. Under the correct conditions, a single base pair mismatch can be sufficient to prevent the formation of a

Communicated by P. Langridge

R. Mogg · J. Batley · S. Hanley · D. Edwards · H. O'Sullivan
K.J. Edwards (✉)
Department of Biological Sciences, University of Bristol,
Woodland Road, Bristol BS8 1UG. UK
e-mail: keith.edwards@bbsrc.ac.uk
Tel.: +44-1275-549431, Fax: +44-1275-394281

stable probe-target duplex. Hybridization/nonhybridization can then be monitored via a suitable detection system. This two-state system is binary in nature and is therefore ideal for automated scoring (Brookes 1999). ASOs can only be designed when sequence polymorphism exists between two individuals. Suitable sequence polymorphisms may consist of either single-nucleotide polymorphisms (SNPs; Brookes 1999) or insertions/deletions (indels). Characterization of SNPs and indels in humans suggest that when comparing two individuals, one SNP or indel can be found in every kilo base pair of sequence (Li and Sadler 1991). Unfortunately, only a limited amount of work has been carried out to examine the occurrence of SNPs and indels in plants. Bryan et al. (1999) found that the sequence variation present between different wheat RFLP alleles was insufficient to design ASOs. However, Germano and Klein (1999) have shown that SNPs are present in the nuclear and chloroplast DNA of both *Picea rubens* and *Picea mariana*. Moreover, these variations were shown to be capable of genotyping individuals more efficiently than RFLPs. In soybean, Coryell et al. (1999) identified two SNPs in 400 bp of sequence from the nuclear RFLP locus A519-1, whereas de Barros et al. (2000) suggested that the flanking regions around soybean SSRs could represent some of the most hypervariable regions of the genome.

Because SNPs and indels, via ASOs, have the potential to be converted into a quick, cheap, multi-allelic and multi-locus test, they should be in regular use within large-scale genotyping laboratories. Unfortunately, whilst they are in regular use for the detection of certain human genetic diseases (Saiki et al. 1989) they are, as yet, not in regular use for non-human genotyping. The reason for this becomes apparent when one considers the enormous cost of developing SNP and indel-based markers. Work by our group and de Barros et al. (2000) has suggested that the amount of effort required to produce such markers would be considerably reduced if existing molecular markers could be converted. These markers would already have been mapped and therefore could be converted based upon their useful map position. Current RFLP and SSR markers offer such a resource. In our search for sources of sequence polymorphisms we chose to examine the existing maize SSR markers available, within either MaizeDB or Genbank (http://www.agron.missouri.edu/maps.html and http://www.ncbi.nlm.nih.gov/Entrez/).

In this study we have compared the sequences from the flanking region of 52 SSR loci and two gene fragments in 11 diverse temperate maize inbred lines. We have used the sequence polymorphisms identified to both determine the level of sequence variation present at the loci and identify individual SSR-linked sequence-based alleles. In addition, the sequence polymorphisms from 32 loci were used to design 123 ASOs, which were then used in a hybridization-based assay to examine differences between the original 11 inbred lines and a second source of the inbred B73.

Materials and methods

The maize inbred lines T232, CM37, T303, CO159 (obtained from Dr Ben Burr, Brookhaven National laboratory), B14, F2, F7, CO125 (obtained from ICI Seeds) and MO17, B73 and OH43 (obtained from Dr Michael McMullen, University of Columbia, Missouri) were used in this study. A second source of B73 was obtained from ICI Seeds.

DNA preparation and amplification

Genomic DNA was extracted from individual 10-day old etiolated seedlings using the procedure of Edwards et al. (1991). Amplification was carried out using primers from 52 SSR loci and two genic loci as described in Table 1. The primer sequences used for these amplifications can be found on the MaizeDB web site (http://www.agron.missouri.edu/maps.html). Amplifications were carried out in a 25 μl reaction volume containing 25 ng of DNA, 2.5 μl of 10 × PCR reaction buffer (Perkin-Elmer), 100 ng of both the forward and reverse SSR primer, 200 μM of each dNTP and two units of AmpliTaq Gold (Perkin-Elmer). After an initial hot-start at 94 °C for 7 min, the following PCR cycling parameters were employed: denaturation at 94 °C for 20 s, annealing at 53 °C for 60 s and extension at 72 °C for 60 s. After 40 rounds of amplification, a final extension step was performed at 72 °C for 10 min. All PCR reactions were carried out in a 9600 DNA Thermal Cycler (Perkin-Elmer).

Purification of PCR products from agarose gels

Following amplification, PCR products were purified by electrophoresis and subsequent elution from 1.2% agarose gels (Hanley et al. 2000).

DNA sequencing of PCR products

Gel-purified PCR products were sequenced using the ABI BigDye Terminator cycle sequencing reaction kit (Perkin-Elmer). Sequencing reactions were carried out in a 10 μl volume containing 2.5 μl of the purified PCR product, 100 ng of either the forward or reverse primer and 4 μl of the BigDye sequencing mix. Cycle sequencing was carried out for 25 cycles in a 9600 DNA Thermal Cycler as described in the Perkin-Elmer handbook. Samples were resuspended in 10 μl of 100% formamide and analyzed using an ABI 377 automated DNA sequencer. To obtain an accurate consensus sequence, individual PCR products were sequenced a minimum of two times and up to five times.

Allele sequences from each SSR locus and inbred line were compared using both the Sequencher (GeneCode) and CLUSTALW programmes employing the default settings. Before the CLUSTALW analysis was carried out, the regions containing the simple sequence repeat were removed. In most cases only one side of the flanking region of the SSR allele was used in the CLUSTALW analysis. Sequence variation between inbred lines was expressed as both sequence-based alleles (i.e. different alleles are sequences that do not share the same SSR flanking sequence) and size-based alleles (i.e. different alleles are fragments that do not share the same size amplification product as judged by denaturing polyacrylamide gel-electrophoresis). It should be noted that in our comparison of the various sequences we have avoided the use of the term ‘haplotype’ to describe sequence-based alleles as, according to the currently accepted definition, this is an incorrect use of the term (Brown 1999). Manual comparisons of the sequence variations were used to design 20-mer ASOs. ASOs were designed to include the maximum number of base pair mismatches between the different sequence-based alleles. In designing the ASOs, no account was made of the GC content of the oligonucleotide.

ASO array preparation and hybridization

ASOs containing a 5′ amino group linked to a 12 carbon chain 'linker group' attached to a 10- mer poly dT followed 3′ by the allele-specific sequence, were custom synthesized by Sigma-Genosys and re-suspended to a concentration of 1 mg/ml in 5 × SSC. ASOs were spotted onto silylated microscope slides (Telechem) using a hand-held arrayer. Silylated glass slides contain reactive aldehyde groups, which covalently bind amino-linked nucleic acids to their surface. After the spotting operation was complete, the slides were re-hydrated in a humid chamber for 4 h and allowed to air dry. Unbound ASOs were removed from the slide surface by washing in 1% SDS for 5 min.

Microscope slides with attached ASOs were pre-hybridized for 1 h, with constant shaking, in 6 × SSC, 0.1% SDS, 5 × Denhardt's solution, 50 mM of $NaPO_4$ pH 7.0 and 1 mg/ml of salmon sperm DNA. Labelling of the gel-purified fragments with α-^{33}PdCTP (Amersham) was carried out using the Ready-to-Go reaction kits supplied by Pharmacia Biosystems. Only 1 µl of the 1:10 diluted gel fragment (DNA unquantified) was needed for probe labelling. Following incubation at 37 °C for 1 h, the probe was phenol-chloroform extracted, denatured and added to the hybridization solution. The hybridization solution was identical to the pre-hybridization solution. Hybridization was carried out in custom-made chambers (VH Biolabs) at 50 °C for 12 h.

Following hybridization, slides were washed twice at room temperature for 10 min in 6 × SSC, 1% SDS. This was followed by two TMAC (3 M Tri-methyl-ammonium-chloride, 50 mM TrisHCl pH 7.5, 2 mM EDTA, 1% SDS) washes at 65 °C for 15 min. Two final washes in 1 × SSC were carried out at room temperature for 10 min. The resulting hybridization pattern was detected by autoradiography and scored independently by two researchers. In all cases, hybridization was determined to have occurred if the intensity of hybridization (as measured by densitometry) to a specific oligonucleotide was greater than 10-fold higher than the background for the slide as a whole. The degree of similarity between the hybridization patterns of the inbred lines used was calculated using UPGMA (unweighted pair-group method using arithmetic averages) clustering analysis with distance matrices calculated using restdist (PHYLIP). The consensus tree's robustness was measured by the bootstrap method (Felsenstein 1985) using 100 data sets.

Results and discussion

Flanking sequence characterisation

Previous work by our group and others has suggested that the regions that flank SSRs contain relatively large numbers of sequence variations, which consist of both SNPs and indels (Grimaldi and Crouau Roy 1997; Mogg et al. 1999; de Barros et al. 2000). To investigate this further, we amplified and sequenced DNA from 52 SSR linked loci and two loci unlinked to SSRs, in 11 maize inbred lines. In all cases, each of the amplified products were sequenced up to five times and the results compared using the Sequencher and CLUSTALW programs. In most cases the position of the primers used in the amplification procedure allowed us to determine the sequence from only one side of the SSR repeat. In addition, the primary SSR motif was removed from the sequences before sequence alignment. An example of a typical sequence comparison is provided in Fig. 1. In this example, locus MMC0071 was found to have one null allele (inbred line CO125). Sequence alignment of the ten remaining sequences indicated that the region contained a total of six SNPs and four indels (not including the primary GA SSR motif) within a total of 2,120 bp of sequence. Therefore, within the sequences generated at this locus, there is a sequence polymorphism for every 212 bp of primary sequence. The total sequence variation at the MMC0071 locus could be accounted for within three distinct sequence-based alleles, with lines T232, B14 and B73 representing sequence-based allele one, lines CM37, CO159, F2 and F7 representing sequence-based allele two and lines T303, M017 and OH43 representing sequence-based allele three. We were able to generate a total of 530 out of a possible maximum of 594 sequences for the 54 loci (89.2%). Within these sequences we found a total of 324 sequence-based polymorphisms (SNPs and indels) or an average of six per loci (Table 1). These sequence polymorphisms consisted of 218 SNPs (representing an average of just over four per locus) and 106 indels (representing an average of just under two per locus). Four of the 54 loci examined (MMC0551, UMC1004, EST600.967 and EST619.213) had no sequence polymorphism. For 15 of the loci (10.8% of sequences), we were unable to confirm the exact consensus sequence of at least one inbred line from five sequencing runs. In these cases the locus-specific sequence information from those inbreds was not analyzed further.

Null alleles (defined as an inability to amplify a specific PCR product from a specific inbred) appeared to be present in at least one inbred line for 18 of the 54 loci (33.3% of all loci or 7.2% of the total number of sequences possible). In each case, when a specific primer set failed to amplify a product, the amplification was repeated with fresh genomic DNA from that inbred before a null allele was considered to be present at that locus. In one case (UMC1007), null alleles were observed in six of the 11 inbred lines, whilst five null alleles were observed at both the UMC1004 and EST619.213 loci. Null alleles are thought to occur due to sequence variation at one or both priming sites (Alexander et al. 1996). Although a figure of 33.3% of loci having null alleles may be regarded as relatively high, high levels of null alleles have been reported previously for animal SSRs. For instance, Alexander et al. (1996) reported that 50 out of 400 (12.5%) porcine SSRs exhibited null alleles. The authors suggested that this number of null alleles indicated that the sequences, which flank SSRs contained a relatively high level of sequence polymorphism. They suggested that for sequence variation at the primer binding site to disrupt PCR amplification the variant nucleotides are likely to be located within five nucleotides at the 3′ end of the primer. Thus, each primer pair assays ten nucleotides for polymorphisms. If this hypothesis is correct, using the information that 7.2% of the total sequences fail to amplify a product, we can predict that for these sequences, there is a minimum of one sequence polymorphism in the ten bases covered by both the forward and reverse primers (2 × 5 bp). In the work described here, this would suggest a minimum polymor-

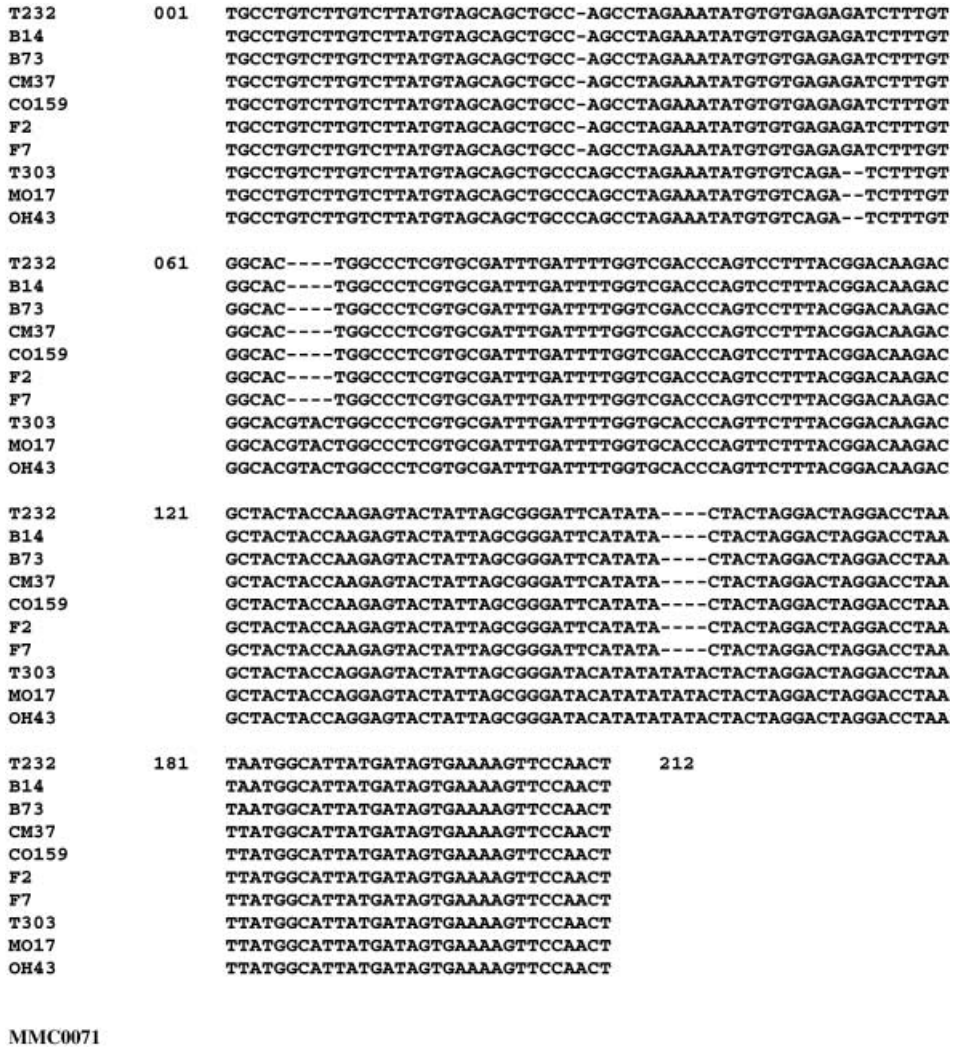


Fig. 1 Sequence alignment of SSR locus MMC0071. The forward primer sequence is not included, but is situated 5′ to base number one. The GA repeat motif is not presented, but is situated 3′ to base number 212. The SNPs and indels are highlighted in *bold*. The sequence for the inbred line C0125 is not included as this line has a null allele at this locus. Inbred lines T232, B14 and B73 had identical sequences (sequence-based allele one), as did inbred lines CM37, C0159, F2 and F7 (sequence-based allele two) and inbred lines T303, M017 and OH43 (sequence-based allele three)

phism rate of one base per 139 bp. Although only a minimum value, this figure is similar to that observed by Alexander et al. (1996), which suggested a sequence polymorphism rate of one polymorphism per 80 base pairs.

Of the 52 SSR linked loci used in this study, 24 were derived for sections of the genome associated with transcribed regions and 28 were thought not to be associated with transcribed regions (Table 1). However, it should be noted that in no case did we sequence beyond the primer sequences and it is therefore possible that some of the "non-EST-linked" sequences are linked to transcribed regions. Six of the 24 (25%) EST linked SSRs had null alleles, whereas 12 of the 28 (43%) non-EST linked SSRs had null alleles. This result suggests that the flanking regions of SSRs that are not linked to transcribed regions, might have a higher rate of polymorphism than those associated with transcribed regions. A comparison of the number of SNPs and indels present per locus in the non-EST linked and EST linked SSRs suggested that there was no significant difference between the two. For instance, the average number of SNPs per locus was 4.08 for non-EST linked SSRs versus 3.89 for EST linked SSRs whereas the average number of indels per locus was 2.08 for non-EST linked SSRs versus 1.85 for EST linked SSRs. However, when corrected for the fact that the average non-EST linked SSR locus used in our study covered 212 bp, compared to 381 bp for the average EST linked SSR locus, the rate of polymorphism per base pair for non-EST linked SSRs was found to be one per 298 bp (range of 78–967 bp) compared to an average of one per 437 bp (range of 68–1,380 bp) for EST linked SSRs (Table 1). In both cases, the two SSRs that did not show any sequence polymorphism across the 11 inbreds were not included in the calculation. Although only two genic regions, not linked to SSRs, were included in our experiments, the average rate of polymorphism for these was found to be one per 229 bp. However, these two genic regions were included in our original experiments because our previous work (K.J.E., unpublished) had suggested that they both had a relatively high rate of sequence polymorphism. The rate of polymorphism for maize SSR flanking regions suggested above and in Table 1 are lower than those determined by the method of Alexander et al. (1996). We believe that this difference is due to the manner in which we have classified

Table 1 Summary of SSR locus information used in this study

Locus	Repeat type	Map location	Source[a]	No. of poor sequences[b]	No. of null alleles	No. of SNPs	No. of indels	No. of base pairs per polymorphism	No. of sequence-based alleles
Non-EST SSRs									
MMC0063	CA	2.00	Maizedb	0	0	12	1	152 bp	2
MMC0071	GA	3.05	Maizedb	0	1	6	4	212 bp	3
UMC128	GA	1.07	Maizedb	1	0	5	6	106 bp	5
MMC0132	GA	3.04	Maizedb	0	1	2	3	224 bp	5
MMC0201	GA	Unknown	Maizedb	0	1	0	2	510 bp	3
MMC0211	CA	unknown	Maizedb	0	0	7	6	88 bp	6
MMC0241	CA	6.05	Maizedb	1	0	6	2	230 bp	4
MMC0261	GA	5.02	Maizedb	1	2	9	1	78 bp	2
MMC0271	GA	2.07	Maizedb	1	0	3	1	193 bp	2
MMC0282	CA	5.05	Maizedb	0	0	3	0	157 bp	3
MMC0321	GA	4.08	Maizedb	2	1	0	3	480 bp	3
MMC0351	GA	5.03	Maizedb	1	2	2	0	178 bp	3
MMC0371	GA	4.06	Maizedb	0	0	2	1	337 bp	2
MMC0381	GA	2.09	Maizedb	0	0	5	0	139 bp	4
MMC0401	GA	2.05	Maizedb	2	0	2	0	357 bp	3
MMC0431	GA	Unknown	Maizedb	0	1	5	4	225 bp	7
MMC0461	GA	Unknown	Maizedb	2	0	4	1	237 bp	3
MMC0471	GA	4.04	Maizedb	0	0	1	1	616 bp	2
MMC0491	GA	Unknown	Maizedb	2	2	2	1	219 bp	3
MMC0501	GA	10.02	Maizedb	0	1	1	0	600 bp	2
MMC0511	GA	Unknown	Maizedb	0	0	0	0	N/A	1
UMC1004	CA	2.05	GB:G42328	0	5	0	0	N/A	1
UMC1007	GA	2.04	GB:G42322	0	6	1	0	965 bp	2
UMC59	GA	3.03	GB:G10853	0	2	2	2	405 bp	5
UMC126	GA	2.03	GB:G10810	0	0	3	4	364 bp	5
UMC1025	GA	3.04	GB:G42323	0	0	7	0	248 bp	4
UMC1027	GA	3.06	GB:G42325	0	0	4	7	275 bp	7
UMC1028	GA	2.05	GB:G42612	0	0	8	1	176 bp	6
EST SSRs									
UMC1003	AAAT	2.04	GB:AF080567	0	0	2	1	795 bp	3
UMC1010	CA	3.09	GB:U66105	1	0	5	2	457 bp	8
UMC1016	GA	7.02	GB:U81960	0	0	4	3	282 bp	7
UMC1022	CA	4.01	GB:X76713	0	0	5	4	217 bp	6
EST491.313	GA	Unknown	GB:AI795636	0	2	8	5	68 bp	7
EST600.908	PolyA	Unknown	GB:AI600908	0	4	3	0	319 bp	2
EST600.967	CA	Unknown	GB:AI692055	0	0	0	0	N/A	1
EST603.742	CA	Unknown	GB:AI603742	0	0	2	1	634 bp	2
EST619.213	PolyA	Unknown	GB:AI967267	0	5	0	0	N/A	1
EST621.450	CA	Unknown	GB:AI795636	0	0	18	2	133 bp	7
EST622.008	PolyA	Unknown	GB:AI622008	0	0	4	3	161 bp	5
EST649.727	AT	Unknown	GB:AI461485	0	1	0	3	440 bp	3
EST649.864	GT/GA	Unknown	GB:AI649864	1	0	2	1	366 bp	5
UMC1170	GA	9.02	GB:AI649893	0	0	7	6	152 bp	6
EST649.899	CAA	Unknown	GB:AI855242	3	0	2	3	328 bp	3
EST665.028	CAA	Unknown	GB:AI833733	0	0	2	0	1,380 bp	2
EST668.131	AT	Unknown	GB:AI668131	0	0	4	1	997 bp	2
EST670.332	GA	Unknown	GB:AI712111	1	0	3	0	766 bp	3
EST670.625	CA	Unknown	GB:AI712273	0	0	6	6	302 bp	6
UMC1127	CA	6.07	GB:AI677270	1	0	2	3	442 bp	3
EST612 250	TAC	10.01	GB:AI612250	3	0	3	2	371 bp	4
EST665.695	TA	7.0	GB:AI665695	0	0	19	4	199 bp	5
EST714.928	GAT	9.07	GB:AI714928	0	0	5	2	469 bp	7
Waxy	TACA	9.03	GB:M24258	0	4	3	2	336 bp	2
Non SSRs									
RAD5IB	None	Unknown	GB:AF079429	0	3	6	2	224 bp	3
GSTI	PolyA	8.09	GB:MI6900	0	0	5	2	235 bp	4

[a] Sequences designated GB were derived from the accession number within Genbank
[b] Poor sequences were defined as sequences that contained ambiguities after five separate sequencing reactions

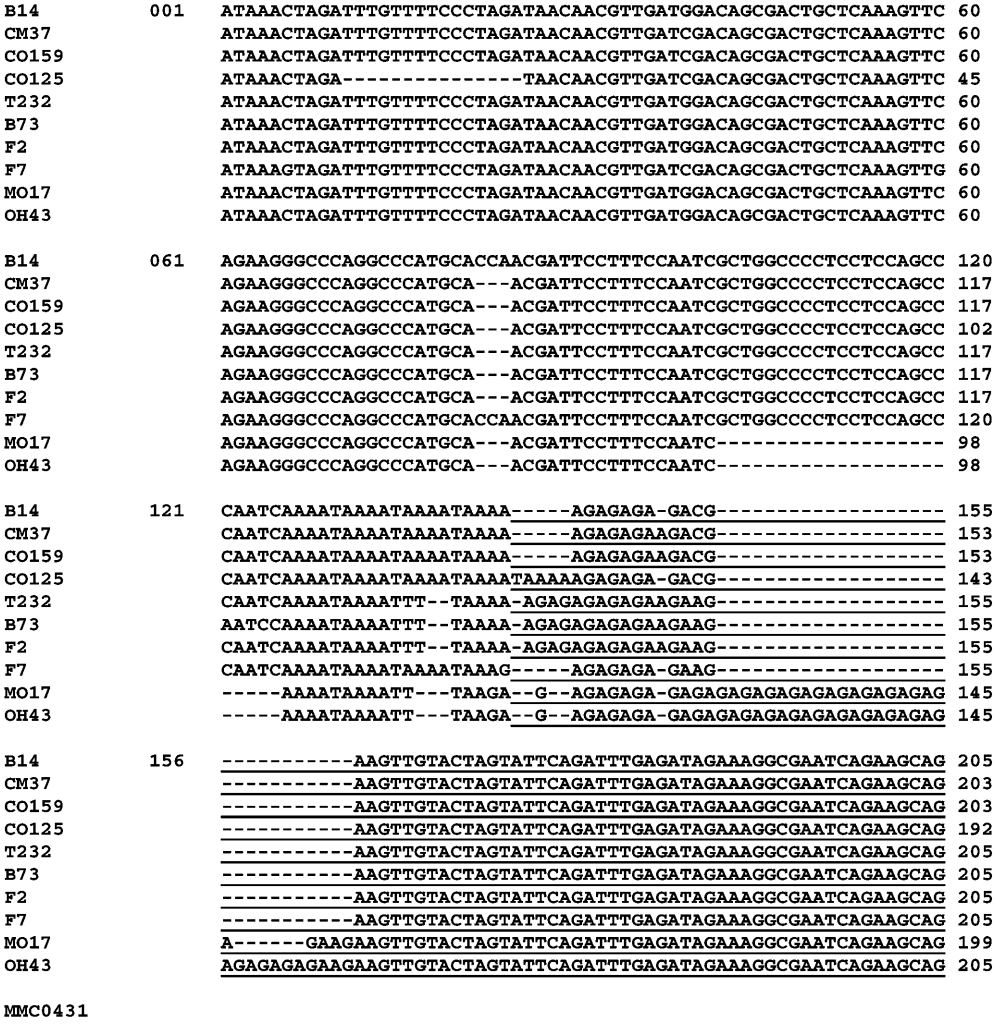


Fig. 2 Sequence alignment of SSR locus MMCO431. The forward and reverse primer sequences are not included, these are situated 5′ to base 001 and 3′ to bases 192–205 respectively. In this sequence alignment the GA repeat motif and 3′ flanking region are *underlined*. The *underlined sequences* were not used in the assessment of either SNPs or indels, but they were used to assess size-based alleles. The sequence for the inbred line T303 is not included as this line has a null allele at this locus. The SNPs and indels used to identify sequence-based alleles are in bold

the various sequence polymorphisms. The maize sequences used here contain a significant number of indels unlike the SSRs described by Alexander et al. (1996). In our study we have classified a single indel as being equivalent to a single nucleotide polymorphism irrespective of the number of bases covered. For instance, locus MMC0071 contains six SNPs and four indels; however, collectively the indels cover an area of sequence equivalent to 11 SNPs (Fig. 1). Given this scenario it is not surprising that the number of sequence polymorphisms suggested by experimentation is significantly higher than that suggested by simply cataloguing the total number of sequence variations. We have chosen this method to classify the observed indels because it is apparent from the various sequence alignments that the bases covered by a single indel cannot be considered as being independent. We believe this to be the case because we see no evidence for individual bases within an indel being inserted or deleted; however, we do see numerous cases where the bases are deleted or inserted as blocks of sequence covering anything up to 25 bases. Interestingly, when indels are present within the flanking sequences they almost always occur in, or next to, short stretches of repeat motifs, which are independent from the main SSRs repeat. For example, of the four indels contained within the sequences derived from locus MMC0071, two are within minor repeats. This observation suggests that the majority of indels are produced by a mechanism similar to that responsible for the variations observed in the major SSR repeat cluster. This observation also suggests that some of the allele length polymorphisms seen with SSRs could be due to the presence of indels within the flanking regions rather than changes in the number of repeats at the primary SSRs motifs. If this is correct then it could lead to SSR homoplasy, whereby different (sequence-based) SSR alleles have evolved to be of identical size (Grimaldi and Crouau Roy 1997). In this case such fragments would be scored as being identical when examined by denaturing polyacrylamide-gel electrophoresis. To find out if SSR homoplasy has occurred within maize SSRs, we compared the number of sequence-based alleles with the number of size-based alleles as determined by denaturing polyacrylamide-gel electrophoresis. In each case the same inbred lines were used for the comparison. The results (Table 2) suggest that for the MMC SSRs (which are not linked to ESTs), the num-

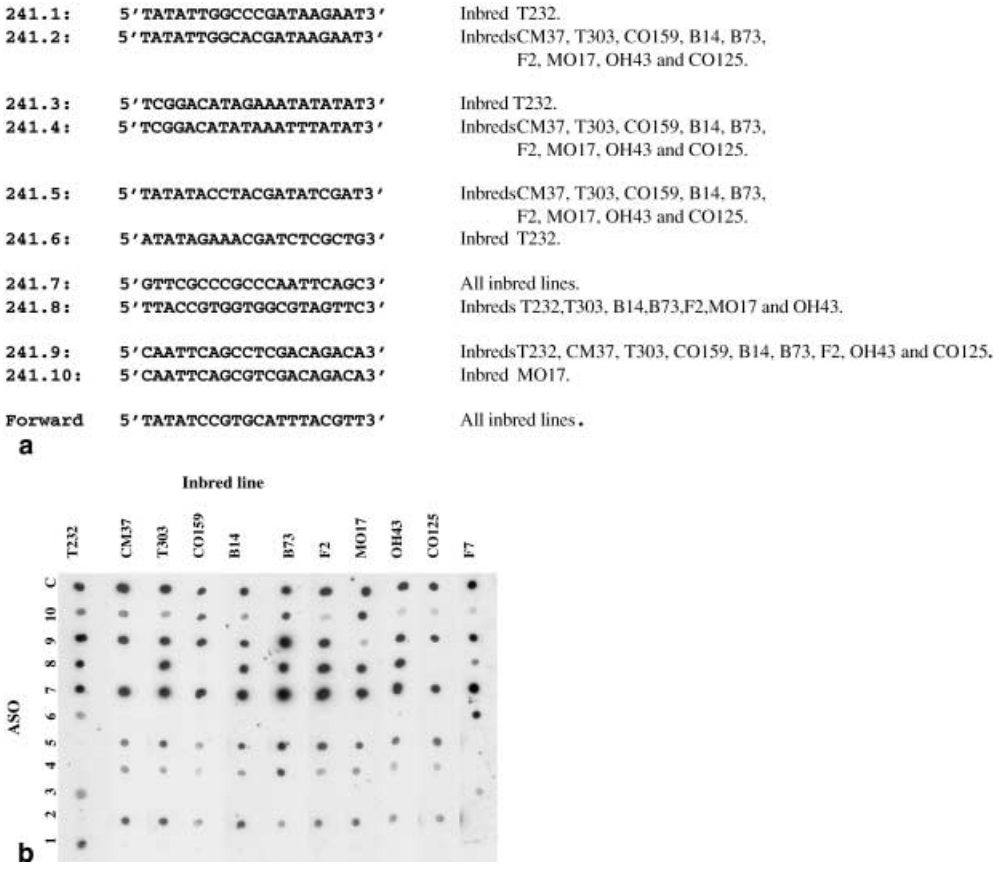


Fig. 3 Allele-specific oligonucleotides designed for SSR locus MMC0241. All oligonucleotides were 20 bases in length. The 10-mer-oligo-dT spacer on the 3′ of each ASO is not shown. The inbred lines, whose MMC0241 PCR product would be expected to hybridise to the ASO are shown. The status of inbred line F7 was unknown at the stage of ASO design. **B** Results of hybridising MMC0241 ^{33}P-labelled PCR products from 11 inbred lines to 11 ASOs bound to glass microscope slides. Note the high level of non-specific hybridization seen with ASO MMC0241-10. ASO C is the forward sequencing primer and was expected to hybridise to all the inbred lines

Table 2 Comparison of the number of sequence-based alleles and size-based alleles for the MMC SSRs

Locus	No. of sequence-based alleles	No. of size-based alleles[a]
MMC0063	2	6
MMC0071	3	4
MMC0132	5	8
MMC0201	3	4
MMC0211	6	5
MMC0241	4	5
MMC0261	2	4
MMC0271	2	6
MMC0282	3	5
MMC0321	3	5
MMC0351	3	6
MMC0371	2	5
MMC0381	4	7
MMC0401	3	9
MMC0431	5	4
MMC0461	3	5
MMC0471	2	8
MMC0491	3	6
MMC0501	2	7
MMC0511	1	3

[a] Allele size was determined using ^{32}P labelled products on 5% denaturing polyacrylamide gel-electrophoresis

ber of sequence-based alleles is less than the number of size-based alleles in 18 of the 20 loci (an average of 3.1 sequence-based alleles compared to an average of 5.5 size-based alleles). However, two loci, MMC0211 and MMC0431, both appear to have a larger number of sequence-based alleles than size-based alleles. In the case of MMC0211, there are six sequence-based alleles compared to five size-based alleles, and in the case of MMC0431 there are seven sequence-based alleles compared to four size-based alleles. Examination of the sequence for each locus from the various inbred lines, clearly showed that, in both cases, changes in the number of repeat units was sometimes compensated for by indels within the SSR flanking region. As an example of this, Fig. 2 shows the entire sequence of MMC0431 in ten inbred lines (inbred line T303 had a null allele for this locus). Examination of the sequences show that for some of the inbred lines, expansion/contraction of the main repeat motif (GA) has been compensated for by the presence of indels within the flanking region. Together this has resulted in size-based homoplasy. For MMC0431, inbred line CO125 has one size-based allele (192 bp), MO17 another (199 bp), CM37 and CO159 a third (203 bp) and B14, T232, B73, F2 and F7 a fourth (205 bp). In comparison, the sequenced-based alleles consist of MO17 and OH43 forming one, B14 another, CM37 and CO159 another, CO125 another, T232 and F2 another, B73 another and F7 the seventh. Therefore for

Table 3 ASO sequences, location on the chip shown in Fig. 4 and inbred specificity

ASO	Sequence (5′3′)	Co-ordinates on chip	Suggested inbred specificity[a]
MMC0071.5	TGTGGCACGTACTGGCCCTC	A1	T303,MO17,OH43
MMC0071.6	TTTGTGGCACTGGCCCTCGT	B1	T232,CO159,B14,B73,F2,F7
UMC128.8	CACCGTGTCTGTGTCCATAC	C1	CM37,CO125,
UMC128.9	ACCGTGTCTAATGTGTCCAT	D1	T232,T303,CO159,F2,OH43,F7
UMC128.10	CACATCAACCAGATACGGTG	E1	B14,B73
MMC0241.1	TATATTGGCCCGATAAGAAT	F1	T232
MMC0241.2	TATATTGGCACGATAAGAAT	G1	CM37,T303,CO159,B14,B73,F2,MO17,OH43,CO125
MMC0241.3	TCGGACATATAAATTTATAT	I1	CM37,T303,CO159,B14,B73,F2,MO17,OH43,CO125
MMC0241.5	TATATACCTACGATATCGAT	J1	CM37,T303,CO159,B14,B73,F2,MO17,OH43,CO125
MMC0241.6	ATATAGAAACGATCTCGCTG	K1	T232
MMC0261.3	TTTGGTTGGGTTGAGCCCAG	L1	T232,CM37,T303,B14,B73,MO17,OH43
MMC0261.4	GGCTAGGTTGGGTTGGGCTT	M1	F2,F7
MMC0271.1	ATTGACGTTAGGAGACATAC	N1	CO159,B14,B73,F2,OH43,CO125
MMC0271.2	ATTGACGTTGGGAGACATAC	A2	T232,CM37,T303,MO17,F7
MMC0321.1	TGTTTGACGAAGGGCGAGGA	B2	MO17
MMC0321.2	CGCTTTGAGGGAGCTGGGTG	C2	MO17
MMC0321.9	ACATGACTCATGAGCTGAGC	D2	CM37,CO159
MMC0321.10	CCCCACATGAGCTGAGCATC	E2	T232,B14,B73,F2,OH43,CO125
MMC0371.1	CCCTTCACCCAGTCAGTCGT	F2	T303,B14,B73
MMC0371.2	CACCCACAGTCAGTAACTCA	G2	T232,CM37,CO159,F2,MO17,OH43,CO125,F7
MMC0431.3	AACGTTGATCGACAGCGACT	H2	CM37,CO159,CO125,F7
MMC0431.4	AACGTTGATGGACAGCGACT	I2	T232,B14,B73,F2,MO17,OH43
MMC0431.5	CTCAAAGTTGAGAAGGGCCC	J1	F7
MMC0431.6	CTCAAAGTTCAGAAGGGCCC	K2	T232,CM37,CO159,B14,B73,F2,MO17,OH43,CO125
MMC0431.7	GGCCCATGCAACGATTCCTT	L2	T232,CM37,CO159,B73,F2,MO17,OH43,CO125
MMC0431.8	CCCATGCACCAACGATTCCT	M2	B14,F7
MMC0431.9	TGGCCCCTCCTCCAGCCCAA	N2	T232,CM37,CO159,B14,B73,F2,CO125,F7
MMC0431.10	CTTTCCAATCAAAATAAAAT	A3	MO17,OH43
MMC0461.8	AAAGCAAAGTGTGCTTGTGT	B3	F2
MMC0461.9	AAAGCTAAGTGTGCATGTGT	C3	T232,CM37,T303,CO159,B14,B73,MO17,OH43,CO12,F7
MMC0491.1	GGAGAAAAGTGTGGTGTCAA	D3	CM37,B73,F2
MMC0491.2	GGAGAAAATTATGGTGTCAA	E3	T232,T303,OH43,CO125
MMC0491.3	GGAGAAAATTGTGGTGTCAA	F3	F7
UMC1003.3	AAGAGAAGGCCATCGAATAA	G3	T232,T303,B14,B73,MO17,OH43,CO125
UMC1003.4	GAGAAGGCGGCCATCGAATA	H3	CM37,CO159,F2,F7
UMC1003.5	TTGAATAAGACGTTGCCCAA	I3	T232,CM37,T303,CO159,B14,B73,F2,MO17,F7
UMC1003.6	TTGAATAAGGCGTTGCCCAA	J3	OH43,CO125
UMC1010.1	CATGGATATGCATGGATGTG	K3	CM37,OH43
UMC1010.2	TTCATGGATATCGATGTGTA	L3	T232,T303,CO159,F2,MO17,CO125,F7
UMC1010.3	TCGATCGACCAACCATTCCG	M3	T232,CM37,CO159,B14,F2,OH43,CO125
UMC1010.4	TCGATCGATCGACCATTCCG	N3	T303,MO17,F7
UMC1012.7	TGAGTGCCAAGGTTCCGTTC	A4	T232,T303,B14,F2,CO125,F7
UMC1012.8	AGTGCCAACAAGGTTCCGTT	B4	CO159,B73,MO17,OH43
UMC1016.9	CTCTATTAATTATAGCTCCC	C4	T232,CM37,T303,B73,F2,OH43,CO125,F7
UMC1016.10	TGCTCTATTATAGCTCCCAA	D4	CO159,MO17
UMC1016.11	TGCTCTATTATATAGCTCCC	E4	B14
UMC1019.1	AGTGGTTACAGACGTACTCC	F4	T232,CM37,T303,B14,B73,F2,MO17,OH43,CO125,F7
UMC1019.2	GGTTATTAGAGACGACGTAC	G4	CO159
UMC1019.3	CGGCCAACAGCTAACCATGC	H4	T232,CO159,B14,B73,F2,MO17,OH43,CO125,F7
UMC1019.4	CGGCCAACACCTAACCATGC	I4	CM37,T303
UMC1022.9	TGACAAGCCGGCTACTAGCT	J4	T232,T303,B14,B73,F2,MO17,OH43,CO125,F7
UMC1022.10	ACAAGCTAAGCCGGCTAGCT	K4	CM37,CO159
UMC1025.8	AGTAATCGGTGGCTTGCGCT	L4	T232,T303,CO159,B14,B73,F2,MO17,OH43,CO125,F7
UMC1025.9	AGTAATCGATGTCTTGCGCT	M4	CM37
UMC1027.3	GCTCAGCCTTAGCAATGGTG	N4	CM37,CO125
UMC1027.4	GCTCAGCCTCAGCAATGGTG	A5	T232,T303,CO159,B14,B73,F2,MO17,OH43,F7
UMC1027.5	TTATCTAGTAGTGTGGCGGA	B5	T232,CM37,T303,CO159,B14,MO17,OH43,CO125,F7
UMC1027.6	TCTAGTACTAGTAGTGTGGC	C5	B73,F2
UMC1027.9	AGCAAAGGCGGAGTGTATAT	D5	CM37,T303,CO159,B73,F2,CO125
UMC1027.10	CAAAGGCGGAGGGAGTGTAT	E5	T232,B14,MO17,OH43,F7
UMC1027.15	CGGAGCAGCTAGCAGAGCTA	F5	T232,CM37,CO159,B14,MO17,OH43,CO125,F7
UMC1027.16	AGCAGCTACTAGCAGAGCTA	G5	T303
UMC1027.17	AGTCGGAGCAGCTAGCGGGG	H5	B73,F2
UMC1028.4	CTGCTTGTTTCACGTGATGC	I5	T303,B14,MO17,OH43
UMC1028.5	CTGCTTGCTTCATGTGATGC	J5	T232,CM37,CO159,B73,F2,CO125,F7
UMC1028.6	TGCATGGAACTGCACCTGAC	K5	F2,CO125,F7
UMC1028.7	GCATGGAACACTGCACCTGA	L5	T232,CM37,T303,CO159,B14,B73,F2,MO17,OH43

Table 3 (continued)

ASO	Sequence (5′3′)	Co-ordinates on chip	Suggested inbred specificity[a]
EST491.313.17	CTCTAGGCGCAGTGACAAGA	M5	T232,OH43
EST491.313.18	AGATAGGTACGGTGACAAGA	N5	T303,B14,MO17,CO125
EST491.313.19	TCTACATAGGTGACAAGATG	A6	CM37,B73,F7
EST621.450.1	AATTTCTACATGGAAAAGGT	B6	T303,F2,OH43,CO125,F7
EST621.450.2	AATTCATGCATGGAAAGGGT	C6	T232,B73,MO17
EST621.450.3	AATTCATGCTTGGAAACGGT	D6	CM37,CO159,B14
EST621.450.23	ACATTTGCCAGATTAACAGA	E6	T303,OH43,CO125,F7
EST621.450.24	ACATTTGCCGGATTAACAGA	F6	T232,B73,F2,MO17
EST621.450.25	ACATTTGCCGAATAACAGAA	G6	CM37,CO159,B14
EST622.008.6	GTTTCTTCCATTATCAAAAA	H6	CM37,B14,OH43,CO125
EST622.008.7	GTTTCTTACATTATCAACAA	I6	T232,T303,B73,F2,MO17
EST622.008.8	GCATCTTCCATTATCAAAAA	J6	CO159,F7
EST649.727.3	TGCTTGCCTAGCTGCCTGTA	K6	CO159,B73,CO125
EST649.727.4	TGCTTGCCTGCCTGTAACGA	L6	T232,CM37,T303,B14,F2,MO17,F7
EST649.864.1	ATGATCGATGGCTACTTGTC	M6	T232,T303,CO159,B14,F2,CO125
EST649.864.2	ATGATCGATGACTACTTGTC	N6	CM37,OH43
EST649.864.3	ATGATCGATTGCTACTTGTC	A7	MO17,F7
EST649.893.1	GGATAGTATAAAATTGCACT	B7	CM37
EST649.893.2	CCAGCGACAAATAAAAAGAA	C7	CM37
EST649.893.10	AAAACGTCACATCGTCGACA	D7	F7
EST649.893.11	AAAACGTCATAACTTCGACA	E7	CO159,B14,F2,OH43,CO125
EST649.893.12	AAAACGTCATATCGTCGACA	F7	T232,T303,B73,MO17
EST649.893.25	GCCCGCTTGCTCCAAGACTT	G7	F2
EST649.893.26	GCCCGCTTGTTCCAAGACTT	H7	T232,T303,CO159,B14,B73,MO17,OH43,CO125,F7
EST668.131.1	GGACATCACGGCCGAGGACG	I7	T232,T303,B14,OH43,CO125,F7
EST668.131.2	GGACATCACAGCCGAGGACG	J7	CM37,CO159,B73,F2,MO17
EST668.131.3	GTTCGTCGGAAGCGGCCTCC	K7	T303
EST668.131.4	GTTCGTCGGCAGCGGCCTCC	L7	T232,CM37,CO159,B14,B73,F2,MO17,OH43,CO125,F7
EST668.131.5	ATTAATGGTCGTGATCTGAT	M7	T303,B73
EST668.131.6	ATTAATGGTGGTGATCTGAT	N7	T232,CM37,CO159,B14,F2,MO17,OH43,CO125,F7
EST670.332.3	CCATGAAGGTACGGGCTTCA	A8	T232,T303,B14,B73,F2,MO17,OH43,CO125,F7
EST670.332.4	CCATGAAAGTACGGGCTTCA	B8	CM37
EST670.625.4	CTGACCGTTGTGTGCTGCAT	C8	CM37,T303,B14,B73,F2,MO17,OH43,CO125,F7
EST670.625.5	GACCGTTGCATTGTGTGCAT	D8	T232,CO159
RAD51B.1	AACAGCTACTGGTCCGGTCT	E8	F7
RAD51B.2	AACAGCTACAGGTCCGGTCT	F8	T232,T303,CO159,B14,F2,MO17,OH43,CO125
EST612.250.7	GCGGTGCTGGTAGTAGTACT	G8	T232,T303,B14,F2,MO17,OH43,CO125
EST612.250.8	GCGGTGCTGTGGTAGTAGTA	H8	CO159
EST665.695.3	TAGTACAAGGTTCGACACGA	I8	B14,F2,MO17
EST665.695.4	TAGTACAAGATTCGACACGA	J8	T232,CM37,T303,CO159,B73,OH43,CO125,F7
EST665.695.7	AGGGAGGGGGTTATATGTGT	K8	OH43
EST665.695.8	AGGGAGGGGCCTATATGTGT	L8	B14,F2,MO17
EST665.695.9	AGGGAGGGGGCTATATGTGT	M8	T232,CM37,T303,CO159,B73,CO125,F7
EST665.695.22	TCGCTGCATACGTGTCTATG	N8	T232,B73,F7
EST665.695.23	ACGTTGCTACATGTCTCTAA	A9	CM37,T303,CO159,B14,F2,MO17,OH43,CO125
EST714.928.8	CGCATGATGATACAACGAAA	B9	CO159,B14,B73,MO17,OH43,CO125
EST714.928.9	TACGCATGATACAACGAAAG	C9	T232,CM37,T303,F2,F7
EST714.928.12	GGTTATTTTGGAGCTGCCAC	D9	T232,CM37,T303,CO159,B14,MO17,OH43,CO125,F7
EST714.928.13	GGTTATTTTAGAGCTGCCAC	E9	B73,F2
GST.3	CCCAAGCATAGGACTGATGA	F9	T232,CO159,B14,B73,OH43,CO125
GST.4	CCCAAGCATCGGACTGATGA	G9	CM37,T303,F2,MO17,F7
GST.12	GAAAGCAACGTCATTAGTAG	H9	T232,CO159,B14,B73,F2,MO17,OH43,CO125,F7
GST.13	GAAAGCAACATCATTAGTAG	I9	CM37,T303
WAXY.3	CACGACGTTGCACTGGGAAG	J9	T232,F2,MO17,OH43,CO125,F7
WAXY.4	CACGACGTTACACTGGGAAG	K9	CM37,B14

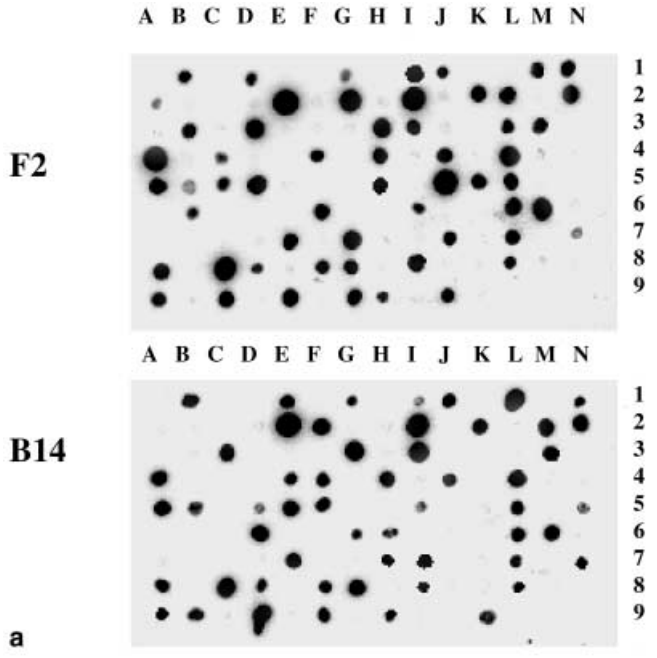
[a] Inbred specificity was determined by examining the sequence alignments

this locus, size-based calling of the alleles would have incorrectly identified B14, T232, B73 and F7 as having the same allele. Overall our observations suggest that approximately 10% of maize SSRs show allele-size homoplasy and hence will give rise to non-identical alleles being scored as identical as judged by co-migration during gel electrophoresis.

Conversion of the SNPs and indels to an ASO-based assay

One of the main purposes of this study was to find a convenient source of sequence polymorphisms, which might be useful for genotyping maize. To assess if the sequence polymorphisms found in this study are of use in genotyp-

ing maize inbred lines, we converted the sequence polymorphisms present in the flanking region of locus MMC0241 into nine ASOs and two control oligonucleotides; one derived from the internal sequence and one derived from the MMC0241 forward PCR primer (Fig. 3A). In each case the ASOs were synthesized with a 5′ amino group linked to a 5′ poly dT tail consisting of ten thymine residues. We chose this approach because it has been suggested that such a "spacer" sequence might help to reduce steric hindrance during hybridization (Guo et al. 1994). Following binding to activated glass slides, the ASOs were hybridized to ^{33}P-labelled MMC0241 amplified products derived in turn from each of the 11 inbred lines. Following overnight hybridization the slides were washed with a wash buffer containing TMAC. The presence of 3 M TMAC in the wash buffer eliminates the influence that base composition (GC content) has on the melting temperature of oligonucleotide – DNA hybrids (Melchior and von Hippel 1973; Dilella and Woo 1987). In our preliminary studies, we had determined that 65 °C was a suitable temperature for the high-stringency washing of DNA duplexes containing 20-mer oligonucleotides and PCR-derived fragments. The results of the hybridization are shown in Fig. 3B. Using the scoring convention described in the Materials and methods, the nine ASOs detect five unique hybridization patterns; inbred line T232 having one pattern, inbred lines CM37, CO159 and CO125 having another, inbred MO17 another, inbreds T303, B14, B73, F2 and OH43 another, whilst inbred F7 has a fifth. Both the internal and external controls showed hybridization to all of the inbred lines used. The result obtained with the ASO-based hybridization procedure was in partial agreement with the sequence data for this locus, which suggested a total of four sequence-based alleles (sequence-based allele one; T232, sequence-based allele two; CM37, CO159 and CO125, sequence-based allele three; MO17 and sequence-based allele four; T303, B14, B73, F2 and OH43. The discrepancy between the number of sequence-based alleles and ASO-based alleles was due to the inclusion of the inbred line F7 in the ASO-based assay, whereas it was excluded (due to poor sequence data) from the sequence-based analysis. Examination of the ASO-based result shows that the F7 hybridization pattern represents the fifth pattern for this locus. Examination of the hybridization pattern in Fig. 3B suggested that whilst the hybridization pattern follows the expected format for ten of the ASOs, in the case of ASO 241.10 it does not. Although this ASO does show preferential hybridization to the MO17 amplification product, it also consistently shows higher than background hybrid-



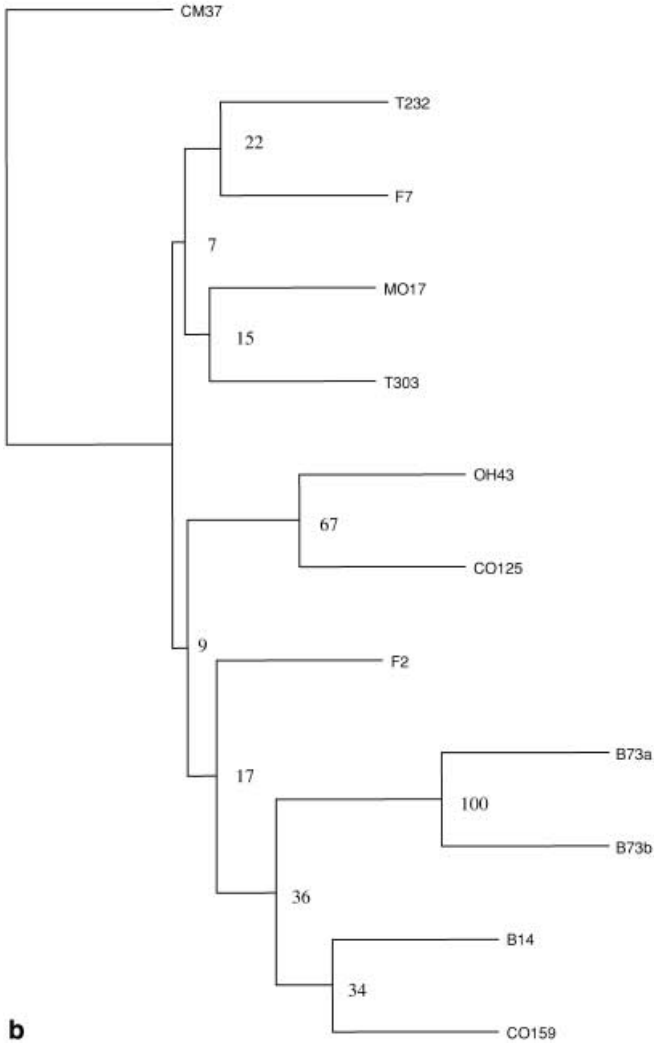


Fig. 4 **A** Results of hybridising 34 locus-specific ^{33}P-labelled PCR products from either inbred lines F2 or B73 to 123 ASOS bound to glass microscope slides. **B** Relationship between inbred lines as judged by ASO-based hybridisation. The degree of similarity between the hybridisation patterns of the 11 inbred lines used and the two different B73 sources used was calculated using UPGMA (unweighted pair-group method using arithmetic averages) clustering analysis with distance matrices calculated using restdist (PHYLIP). The consensus tree was bootstrapped using 100 data sets

ization to all of the other inbred lines. This result was not expected, as the PCR products from the remaining ten lines did not share full sequence homology to ASO 241.10. Further washing at higher stringency (70 °C) did not reduce this non-specific hybridization. The reasons for this inconsistency are unclear; however, this ASO and its counterpart ASO 241.9 share a GC-rich core sequence in which the only polymorphic base is a G to C at position ten. Such polymorphisms may prove unsuitable for such discrimination as described here. Although the result with ASO 241.10 suggest that not all ASOs are suitable for discriminating between sequence polymorphisms, the overall results indicate that by using TMAC in the washing buffers, it is possible to hybridize ASOs with significantly different melting temperatures under the same conditions and produce results which correspond to the known genotype information.

Following our observations with locus MMC0241, 123 20-mer ASOs from 32 SSR-linked loci and two gene-linked loci were designed and synthesized (Table 3). A minimum of two and a maximum of nine ASOs were designed for each locus. Again no attempt was made to match the melting temperatures of the ASOs designed. The 123 ASOs were bound to a single glass microscope slide in the format described in Table 3. Individual microscope slides were hybridized with the combined ^{33}P-labelled PCR products from one inbred, for each of the 34 loci. Following hybridization the slides were washed in the presence of TMAC as described previously. Each hybridization was carried out three times and the results scored as before. A typical autoradiograph for inbreds F2 and B14 is shown in Fig. 4A. The hybridization results are in general agreement with the predicted results (Table 3), confirming the ability of both locus and ASOs to discriminate between large numbers of non-homologous PCR products under identical conditions. However, as described for locus MMC0241 the results also suggested that approximately 5% of the ASOs did not behave as predicted. For instance, in Fig. 4A ASO UMC1027.5 (co-ordinate B5) was not expected to hybridize to PCR products derived from genotype F2 and ASO UMC1027.9 (co-ordinate D5) was not expected to hybridize to PCR products derived from genotype B14. In addition, ASO UMC670.625.5 (co-ordinate D8) was not expected to hybridize to PCR products derived from both genotypes F2 and B14. Of the 123 ASOs used in this study six (4.8%) appeared to have an aberrant hybridization pattern. The reasons for the aberrant hybridization patterns are unclear; however, in all cases when an ASO did have an aberrant hybridization pattern, this pattern was consistent within and between genotypes and hybridization experiments, thus ruling out artifacts. When converted to a binary score ("1" being used to designate hybridization and "0" being used to designate no hybridization) each inbred line generated a unique code of 123 characters. Conversion of genotypes to a binary code could be extremely useful for both genotype identification and purity testing across diverse laboratories as it would provide a common scoring system. To test the robustness of the ASO assay, we also screened the 123 ASOs with PCR products from a second B73 line obtained from ICI Seeds in 1994. The binary code generated from this experiment was compared to the scores generated from both the original B73 line and the other ten inbred lines used via UPGMA, and the results visualized as a dendrogram (Fig. 4B). Our results clearly suggested that whilst the second B73 line is more closely related to the original B73 lines than the remaining ten inbreds, they are not identical. Altogether the two B73 lines differ in their hybridization pattern at seven out of the 123 ASOs. This compares with 34 differences in the hybridization pattern of the first B73 line and line B14. Such a result was not unexpected as the two B73 lines had been obtained from independent sources and, in addition, both B73 lines had been grown (independently) within our department for at least six generations. There had therefore been numerous opportunities for genetic contamination to occur. Interestingly, the level of polymorphism detected by the 123 ASOs was so high that the confidence limit of the UPGMA analysis for all but the B73: B73 comparison was relatively low as judged by bootstrapping (Fig. 4B). This result would suggest that, whilst the genotyping assay described here would be extremely powerful for assessing an inbred line's origin and purity, it would be less useful for examining populations with even greater levels of diversity; for instance, temperate versus tropical maize.

In conclusion, we have developed several hundred SNP and indel markers from 54 maize loci. In the majority of cases these loci are derived from existing SSR markers, for which the primer sequence and map location is known. From the results presented here it appears that such regions of the genome are a valuable source of sequence-based molecular markers. We have also shown that the underlying mechanism which leads to these sequence variations can result in problems with scoring SSR loci via gel-based technology. However, our work has also shown that these markers, when converted to ASOs, can be used to rapidly characterize inbred lines simultaneously for several tens of loci. Such genotyping could be very useful to both probe the origins of agronomically important loci and assess the purity and/or origin of the inbred lines used for breeding purposes.

Acknowledgements Rebecca Mogg, Jacqueline Batley and Helen O'Sullivan were supported by a BBSRC-GAIT award (GAT090151). The authors thank Dr. Alan Archibald of the Roslin Institute for many hours of stimulating conversation on SNP-based technology. All the sequences and information described in this report are freely available, as an EXCEL spreadsheet, for research applications by writing to Keith J. Edwards.

References

Alexander LJ, Rohrer GA, Beattie CW (1996) Cloning and characterization of 414 polymorphic porcine microsatellites. Anim Genet 27:137–148

Barros EG de, Tingey S, Rafalski JA (2000) Sequence characterization of hypervariable regions in the soybean genome: leucine-rich repeats and simple sequence repeats. Genet Mol Biol 23:411–415

Brookes AJ (1999) The essence of SNPs. Gene 234:177–186
Brown TA (1999) Genomics, 1st edn. Bios Scientific publishers, Oxford, UK
Bryan GJ, Stephenson P, Collins A, Kirby J, Smith JB, Gale MD (1999) Low levels of DNA sequence variation among adapted genotypes of hexaploid wheat. Theor Appl Genet 99:192–198
Coryell VH, Jessen H, Schupp JM, Webb D, Keim P (1999) Allele-specific hybridization markers for soybean. Theor Appl Genet 98:690–696
Dilella AG, Woo Savio LC (1987) Hybridisation of genomic DNA to oligonucleotide probes in the presence of tetramethylammonium chloride. Methods Enzymol 152:447–451
Edwards K, Johnstone C, Thompson C (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Res 19:1349
Felsenstein J (1985) Confidence-limits on phylogenies–an approach using the bootstrap. Evolution 39:783–791
Germano J, Klein AS (1999) Species-specific nuclear and chloroplast single nucleotide polymorphisms to distinguish *Picea glauca P-mariana* and *P-rubens*. Theor Appl Genet 99:37–49
Grimaldi MC, Crouau Roy B (1997) Microsatellite allelic homoplasy due to variable flanking sequences. J Mol Evol 44:336–340
Gonen D, VeenstraVanderWeele J, Yang Z, Leventhal BL, Cook EH (1999) High-throughput flourescent CE-SSCP SNP genotyping. Mol Psychiatry 4:339–343
Guo Z, Guilfoyle RA, Thiel AJ, Wang R, Smith LM (1994) Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass support. Nucleic Acids Res 22:5456–5465
Gupta PK, Balyan IS, Sharma PC, Ramesh B (1996) Microsatellites in plants: a new class of molecular markers. Curr Sci 70:45–54
Hanley S, Edwards D, Stevenson D, Haines S, Hegarty M, Schuch W, Edwards KJ (2000) Identification of transposon tagged genes by the random sequencing of *Mutator*-tagged DNA fragments from *Zea mays*. Plant J 22:557–566
Helentjaris T, Slocum M, Wright S, Schaefer A, Nienhuis J (1986) Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphism. Theor Appl Genet 72:761–769
Li W, Sadler LA (1991) Low nucleotide diversity in man. Genetics 129:513–523
Melchior WB, von Hippel PH (1973) Alteration of the relative stability of dA.dT and dG.dC base pairs in DNA. Proc Natl Acad Sci USA 70:298–302
Mogg R, Hanley S, Edwards KJ (1999) Generation of maize allele specific oligonucleotides from the flanking regions of microsatellite markers. Plant and Animal Genome Conference: PI491
Saiki RK, Walsh PS, Levenson CH, Erlich HA (1989) Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. Proc Nat Acad Sci USA 86:6230–6234
Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbour Laboratory, Cold Spring Harbour, New York
Vos P, Hogers R, Bleeker M, Rijans M, Van de Lee T, Hornes M, Frijters A, Pots J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414
Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44:388–396