Theor Appl Genet (2002) 104:399–407 © Springer-Verlag 2002

I. Eujayl · M.E. Sorrells · M. Baum · P. Wolters
W. Powell

Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat

Received: 19 December 2000 / Accepted: 17 April 2001

Abstract Genetic variation present in 64 durum wheat accessions was investigated by using three sources of microsatellite (SSR) markers: EST-derived SSRs (EST-SSRs) and two sources of SSRs isolated from total genomic DNA. Out of 245 SSR primer pairs screened, 22 EST-SSRs and 20 genomic-derived SSRs were polymorphic and used for genotyping. The EST-SSR primers produced high quality markers, but had the lowest level of polymorphism (25%) compared to the other two sources of genomic SSR markers (53%). The 42 SSR markers detected 189 polymorphic alleles with an average number of 4.5 alleles per locus. The coefficient of similarity ranged from 0.28 to 0.70 and the estimates of similarity varied when different sources of SSR markers were used to genotype the accessions. This study showed that EST-derived SSR markers developed in bread wheat are polymorphic in durum wheat when assaying loci of the A and B genomes. A minumum of ten EST-SSRs generated a very low probability of identity (0.36×10^{-12}) indicating that these SSRs have a very high discriminatory power. EST-SSR markers directly sample variation in transcribed regions of the genome, which may enhance their value in marker-assisted selection, comparative genetic analysis and for exploiting wheat genetic resources by providing a more-direct estimate of functional diversity.

Communicated by P. Langridge

I. Eujayl (✉) · M. Baum
ICARDA, Germplasm Program, P.O. Box 5466, Aleppo, Syria

M.E. Sorrells
Department of Plant Breeding and Biometry, 520 Bradfield Hall, Cornell University, Ithaca NY 14853, USA

P. Wolters · W. Powell
DuPont Agriculture Biotechnology, DTP 200, 1 Innovation Way, Newark, DE 19714, USA

Present addresses:
W. Powell, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, Scotland, UK

The S.R. Noble Foundation, 2510 Sam Noble PKWAY, Ardmore, OK, USA

Keywords Durum wheat · EST-SSR · Genotyping · Functional diversity

Introduction

Microsatellites (Tautz and Renz 1984; Tautz 1989), or simple sequence repeat (SSRs)-based molecular markers are now the marker of choice in most areas of plant genetics. The advantages of SSRs are well documented (Powell et al. 1996) and these include: high information content, co-dominant inheritance, reproductibility and locus specificity. This latter attribute is of special importance and relevance when one considers the application of SSR genotyping to germplasm collections, which often requires a long-term commitment to accurate data acquisition from different sources. In addition to their utility for practical applications, the dynamics of microsatellite evolution have been extensively studied, particularly in mammalian genomes (Bowcock et al. 1994; Di Rienzo et al. 1994; Chakraborty et al. 1997). These studies have highlighted the complex mutational patterns operating at SSR loci, emphasising the need to consider opportunities to extract microsatellites from different genomic regions that may be under different selective forces.

Traditional experimental methods for developing SSRs are based on isolating and sequencing clones containing putative SSR tracts, together with designing and testing flanking primers. In order to reduce costs, various pre- and post-cloning procedures (reviewed by Powell et al. 1996) to create genomic libraries enriched for SSRs have been developed. These technical developments have made a significant contribution by generating large numbers of SSRs in crop plants. The practical utility of such markers in germplasm analysis, where data integration and comparison are critical, requires that each SSR is validated for quality and robustness of the amplification product. These approaches may be enhanced by identifying SSRs in coding as well as non-coding regions of plant genomes.

Significant gene-discovery programs in both the public and private sector have been initiated based on single-

pass sequencing of complementary DNAs (cDNAs) derived from populations of messenger RNA (mRNAs) from various tissues. The basic strategies and rationale for such programs are well-documented (Adams et al. 1992) and provide the most-direct route to homology based gene identification. The initial focus of such efforts has been on the production of expressed sequence tags (ESTs) from mammalian genomes (Adams *et al 1992,* Takahashi and Ko 1994), *Caenorhabitis* (Waterston et al. 1992) and various microorganisms (Vassarotti and Goffeau 1992; Ajioka 1998). However, significant emphasis has been recently placed on generating substantial EST databases for plants species such as *Arabidopsis* (Delseney et al. 1997) and crop plants such as grape (Scott et al. 2000), rice (Yamamoto and Sasaki 1997; Cho et al. 2000), and maize (Briggs 1998). In addition, to providing opportunities for gene discovery, such databases may also provide a novel source of microsatellites (SSRs) that are physically associated with coding regions of the genome (EST-derived SSRs). To address this issue we have screened 117,000 wheat ESTs from the DuPont Company database to identify a sample of SSR markers in order to evaluate their utility and informativeness for genotyping wheat.

Durum wheat [*Triticum turgidum* (L.) Thell. Convar. *durum* (Desf.) MK] is a tetraploid wheat species (genomes AABB) which is mainly grown in the Mediterranean region, Canada, USA, Argentina and India. Despite its agricultural significance this species has received relatively little genetic-research attention. A genetic linkage map based on RFLPs (Blanco et al. 1998) has recently been published and genomic microsatellite markers developed for bread wheat have been integrated with the RFLP map (Korzun et al. 1999).

For effective conservation and the use of genetic resources, evaluation of the genetic variation within collections is crucial and could be dramatically enhanced by using molecular genotyping tools. Evaluation of germplasm with SSRs derived from ESTs (EST-SSRs) may enhance the role of genetic markers by assaying variation in transcribed and known-function genes. The overall purpose of this project was to compare the utility of various sources of wheat SSRs to discriminate between durum genotypes that had been previously characterised with RFLPs, morpho-physiological traits, gliadins and coefficient of parentage (Sorrells et al. 1995; Autrique et al. 1996). Furthermore, since the EST-SSR markers represent transcribed regions of the genome, it is also of interest to compare estimates of genetic diversity calculated from both genomic and EST-derived SSR markers.

In this manuscript we report our results on: (1) an evaluation of the ability of EST SSR markers to detect polymorphism between durum wheat genotypes obtained from a world-wide collection and compare their usefulness with SSR markers derived from genomic DNA, and (2) the generation of a set of fluorescentally labelled SSRs that are of high quality for genetically fingerprinting durum wheat cultivars. These SSRs provide a robust set of PCR primers that will allow fingerprinting and the integration of data from the A and B genomes of wheat.

Materials and methods

Genetic material

Sixty-four durum wheat selections, landraces and varieties were obtained from the Genetic Resources Unit of ICARDA in Aleppo, Syria (Table 1). This subset of durum lines originates from differ-

Table 1 Durum accessions genotyped with SSRs together with their country of origin

Genotype	Country	Genotype	Country	Genotype	Country
Hedba	Algeria (1)[a]	Bicre (Bittern/Crane)	Var. (6)	M 21	Morocco (1)
Oued zenati 368	Algeria (2)	Furat 1	Var. (7)	M 3	Morocco (2)
Jennah Khetifa	Algeria (3)	Nile	Var. (8)	M 20	Morocco (3)
Pelissier	Can/USA	Hazar	Var. (9)	M 1090	Morocco (4)
Wakooma	Canada (1)	Om Rabi 14	Var. (10)	Casablanca	Portugal (1)
DT 369	Canada (2)	Karasu	Var. (11)	Santa Marta	Portugal (2)
Kyperounda 1	Cyprus	Gr/Boy	Var. (12)	Entre Largo	Portugal (3)
IC8373	Ethiopia	Sajur	Var. (13)	Corado11639	Portugal (4)
Vatan	ExUSSR	Ain Arous	Var. (14)	Haurani Nawani	Syria (1)
Agathe	France (1)	Sebou	Var. (15)	Huarani 27	Syria (2)
Ardente	France (2)	Belikh	Var. (16)	Akbash	Syria (3)
Neoder	France (3)	Cham1(Frigate,Waha)	Var. (17)	Kishk	Syria (4)
Romanou 2	Greece (1)	Lahn	Var. (18)	Baladia Hamra	Syria (5)
Mavragani Iraklion	Greece (2)	Siliana	Var. (19)	Razzak	Tunisia
Moundros 2	Greece (3)	Heider(Marajawi)	Var. (20)	H.O.FAO25918	Turkey (1)
Atsiki 3	Greece (4)	Iran 1	Iran	Kirmizi	Turkey (2)
Local Iraklion	Greece (5)	Tripolino	Italy (1)	Aric31708	USA (1)
Tensifit 1	Var.[b] (1)	Scorsonera	Italy (2)	N.D Line No.10	USA (2)
Gedifla	Var. (2)	Sicilia Lutri	Italy (3)	Rugby	USA (3)
Chahba 88	Var. (3)	Sentore Cappelli	Italy (4)	Ward	USA (4)
Sabil 1	Var. (4)	Coll 86 NO 42	Jordon (1)		
Loukos 1	Var. (5)	Coll 86 No 44	Jordon (2)		

[a] Figures in parenthesis refer to the genotypes in Fig.1
[b] Improved variety

ent geographic regions and different breeding backgrounds. The wheat variety Chinese spring was used as the DNA and RNA template for the construction of the wheat microsatellite markers. The International Triticeae Mapping Initiative (ITMI) mapping population, Opata 85 x W7984 RILs, was used for genetic-linkage mapping.

Microsatellite markers sources

Genomic SSR primer information was obtained from two sources. The first primer set was obtained from Röder et al. (1995, 1998) from a conventional genomic library and designated as GWM, and the second was obtained from the wheat microsatellite consortium (WMC) provided by Agrogene, S.A., from genomic libraries enriched for wheat SSRs. EST sequences from the DuPont company were subjected to BLAST (Altschul et al. 1990) to identify SSRs. The EST-SSR primers were designed at the flanking regions of the repeats with Primer 3 software (S. Rozen and H.J. Skaletsky; available at http://www-genome.wi.mit.edu/genome_ software/other/primer3.html). The primer-design criteria included a Tm of 60 to 65°C with no greater than a 3°C difference in the Tm between primer pairs. The primers were designed to produce PCR products ranging from 150- to 300-bp in length. The DuPont SSR set was designated as DuPw, and their primer sequence information is given in Table 3. The sixty four durum DNA samples were screened for polymorphism with 137 EST-SSR (DuPw), 68 GWM and 40 WMC microsatellite primers.

SSR marker detection

DNA was isolated from five to eight seedlings (3 to 5 days old) using the DNAzol reagent (Gibco/BRL) and applying the method described by Shure et al. (1983). For PCR analysis of sample DNA, the forward SSRs primer of each pair (genomic or EST) were commercially labelled by the addition of fluorescent phosphoramidite as either 6-carboxyfluorescein (6-Fam), hexachloro-carboxyflourescein (HEX) or NED at the 5′ end of the oligonucleotide (Applied Biosystems/Perkin Elmer). The PCR-amplification conditions were as described by Röder et al. (1998) with some modifications. PCR products of samples (10 µl) were diluted to approximately one-tenth. One microliter of diluted sample was added to 3 µl of loading mixture made up of 0.5 µl of a *GenScan* 500 ROX (6-carboxy–×–rhodamine) internal size standard, 2 µl of de-ionised foramide and 0.5 µl of loading dye. The samples were mixed with 3 µl of loading mixture, denatured at 95°C for 2 min and kept on ice until loading. Approximately 1 µl was loaded on 6% (w/v) denaturing acrylamide:bisacrylamide (19:1) gels. Samples were electrophoresed in 1× TBE buffer at constant power (2000 W) for 2 to 3 h depending on the expected fragment size of the SSR in an automatic DNA sequencer (ABI 377, Applied Biosystems). The ABI 377 is equipped with *GenScan* 3.0 software (Applied Biosystems) for data collection and fragment-size (bp) calculation to two decimals. To assure precision and reproducibility during the polymorphism screening process all polymorphic primers were run at least twice.

Statistical analysis

Electropherograms obtained by *GenScan* 3.0 from the gel images (see Fig. 1) were scored for allele size and presence or absence in each sample. Presence or absence was coded by 1 or 0, respectively, and represented as a binary matrix. The genetic similarity matrix was obtained by the SIMQUAL sub-routine of the NTSYS-pc software statistical package (Rohlf 1990) based on Jaccard's algorithms. The similarity coefficients were used to construct genetic-distance phenograms using the SAHN method based on UPGMA.

Allelic variation is calculated from the frequencies of genotypes at each locus as the diversity index (DI) according to Weir (1990), i.e., $DI = 1 - \Sigma pi^2$, where pi represents the frequency of the ith SSR allele. To account for residual heterogeneity a direct count of heterozygosity was calculated as the number of genotypes, which were heterozygous (more than one band amplified by one pair of primers) at a given locus, divided by the total number of genotypes scored at that locus. The probability of identity (PI) was calculated as: $PI = \sum pi^4 + \sum_{i=1}^{i=n-1} \sum_{j=i+1}^{n} (2pipj)$ according to Paetkau (1995). The PI over a number of loci was calculated as the product of the PI for those loci.

Results

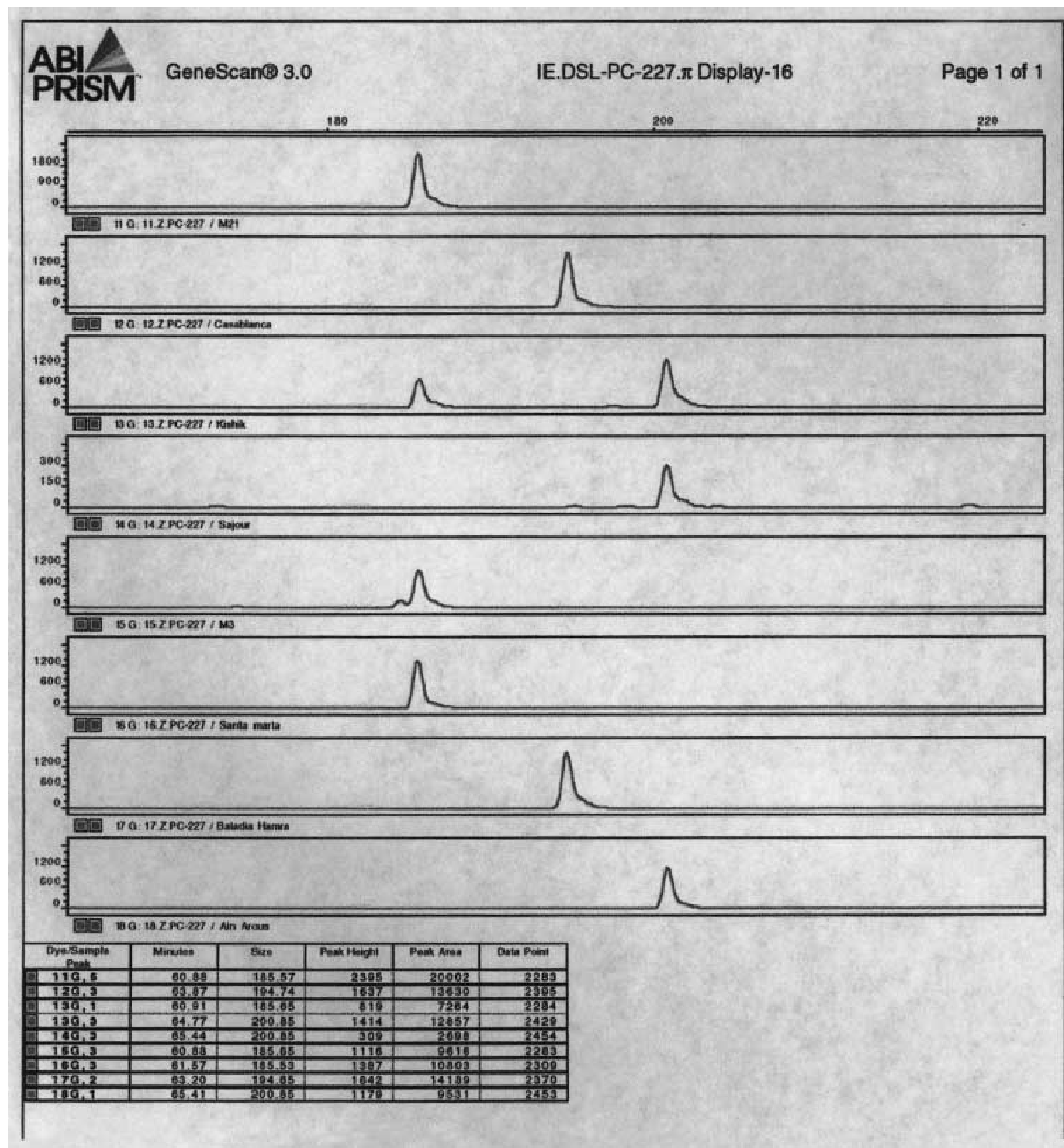
SSR marker polymorphism

The primer sequences of the EST-SSRs, which range between 18 and 25 bp, are presented in Table 3. Based on the screening of 245 pairs of primers for fragment amplification, fragment quality and polymorphism, 42 were identified for fingerprinting the 64 durum genotypes (Table 2). Approximately 25% of the EST-derived SSRs were polymorphic compared to over 50% for the other sources of SSR markers. However, both the GWM and WMC primers had been previously screened for informativeness in hexaploid wheat. An important feature of the EST-derived SSR markers is the high quality fragment-patterns obtained which are devoid of stuttering (Fig 1.). In this example we have an A-genome amplification product generated from primers that amplify a trinucleotide repeat (AAG) that detects a total of six alleles. The clarity of the amplification profiles will certainly ease the application of EST-derived SSR markers in cultivar identification. The 22 selected EST clone sequences were subjected to BLAST; 11 showed convincing homology with gene sequences for which a putative gene function is established (Table 3). The characteristics of the PCR-amplified fragments is presented in Table 4, including the number of repeats, the allele sizes amplified in Chinese Spring, allele size range amplified

Table 2 Number of SSR primers tested, percentage-detecting polymorphism and the average number of alleles detected in the durum accessions studied

Source of SSRs	Number of primers tested	% Polymorphism	Number of loci selected	Number of alleles detected	Average number of alleles
EST-SSR	137	24.8	22	89	4.1
GWM[a]	68	52.9	11	61	5.5
WMC[b]	40	52.5	9	39	4.3
Total	245	35.9	42	189	

[a] Primers from Röder et al. (1998)
[b] Primers from wheat microsatellite consortium (Agrogene)



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
11G, 6	60.88	185.57	2395	20002	2283
12G, 3	63.87	194.74	1637	13630	2395
13G, 1	60.91	185.65	819	7264	2284
13G, 3	64.77	200.85	1414	12857	2429
14G, 3	65.44	200.85	309	2698	2454
15G, 3	60.88	185.65	1116	9616	2283
16G, 3	61.57	185.53	1387	10803	2309
17G, 2	63.20	194.65	1842	14189	2370
18G, 1	65.41	200.85	1179	9531	2453

Fig. 1 Electropherogram obtained from the gel image of primer DuPw227 showing the fragment peaks for eight durum genotypes and information on fragments size

in the durum genotypes and the chromosomal location of the SSR markers based on the ITMI population genetic-linkage map (P. Wolters unpublished data). The number of alleles detected and the diversity index (DI) for each microsatellite marker are also given in Table 4. A total of 189 alleles were identified, 89, 61 and 39 being detected with the EST-SSR, GWM and WMC SSR markers respectively. The most-informative microsatellite marker was GWM169, which detected 14 alleles in the sample of durum genotypes evaluated. Null alleles were detected with three EST-SSRs (DuPw 123, 165, 207).

Genetic relatedness of durum accessions and the discriminatory power of SSRs

The probability of identity for any two genotypes (PI) was calculated for each SSR marker and were quite variable, as were the diversity indices reflecting the range and frequencies of alleles detected by each SSR marker (Table 4). The probability of a chance match between any two genotypes at a given locus in this collection ranged from 0.009 at GWM169 to 0.940 at GWM149. The PI at only ten selected EST-SSR loci (shown by *asterisks* in Table 4) was 0.361×10^{-12}. Therefore, it is possible to identify a small set of EST-SSR markers to uniquely fingerprint durum genotypes. The genetic similarity between genotypes is depicted in the form of a

Table 3 Primer information for the EST-SSRs and information on homology to genes of known function

Locus name[a]	BLAST results		Putative gene function	5′ Primer	3′ Primer
	Score	pLog			
DuPw004	83	5.7	ABA induced protein, PM-19 in *T. aestivum*	GGTCTGGTCGGAGAAGAAGC	TGGGAGCGTACGTTGTATCC
DuPw023	265		Alcohol dehydrogenase in barley 41	TTGCTCCGATGTAATAAGCG	GGCTAAGAACAGACTCATTCAACTG
DuPw038	108	5–10	ADP/ATP translocator (3′ end) in wheat	ATTAGACACGACCAAACGGG	TCAAACAAACAACAGCCAGC
DuPw043	113		Cytochrome P-450 in *Sorghum bicolor* 24	TTTGAACGGAATTTGAGAATTT	AGGGTGTGAACATGGAGGAG
DuPw108	112	4.16	Phosphoenol Pyruvate Carboxylase (ATP)	TGAAGAGTGCGATGTGAAGG	TGTGACAGAAACTACTAACATTGCG
DuPw115	1249	120.4	SS N. A.[b] binding protein, wheat	TGTTTCTTCCTCGCGTAACC	CCTCGAATCTCCCAGTTATCG
DuPw123	198	18.13	Beta-1,3 glucanase 2 in barley	CAACGAGAACCAGAAGACCG	CCCGTTACACTTGGATGCC
DuPw124	141	15.05	CERI-like 3’ gene fragments in *Arabidopsis*	AGCCAAGCCAGTCCAAGC	ACGCGAGAAGGATTATTTGG
DuPw135	–	–	–	CGCTTCTCTTGTCTTGTCCC	CATGGTGAAGACGGTGACG
DuPw165	–	–	–	TAGGTCTCGACAACAAGCCG	TCACCACTTGGAGGTTACTGC
DuPw167	–	–	–	CGGAGCAAGGACGATAGG	CACCACACCAATCAGGAACC
DuPw168	–	–	–	AAGAATAGAATAAGGGTACACGGC	ATTTGTGATTGTAGCCACGG
DuPw173	389	44.24	Liquid transfer protein in barley	AGATGCCGAGTACGTTGAGG	GTGAACGAACAATCAAACGC
DuPw205	–	–	–	ATCCAGATCACACCAAACGG	CTTCCGCTTCATCTTCTTGC
DuPw207	–	–	–	GAGAGTATCAATAAAGCTAGATGCCC	GCATTTGGAAGGAGATGTGG
DuPw210	–	–	–	CGATTTGGATTCTTCCGC	AGAGCCTTTGAAGAGCAGGG
DuPw216	–	–	–	ACAAACCTCTCCCTCTCACG	ATGATGATTCAGCGAGTCGG
DuPw217	–	–	–	CGAATTACACTTCCTTCTTCCG	CGAGCGTGTCTAACAAGTGC
DuPw227	104	4.96	Hypothetical protein in *Arabidopsis*	CATGTTGGGAATTTCTGTCG	CCACGAGCCATGTATCACC
DuPw238	169	14.14	Proline transporter 1, *Arabidopsis*	TTCATAGACGCAACTAGCCG	GACTTTGGTTGTTAAAGGCG
DuPw254	–	–	–	TTAACCATGCAGCAACTTCG	GTGTGTACTAACGGCTACGGC
DuPw398	–	–	–	CTGAGCCCTCTTTGCTATGC	TCGGTGAGATTGAAAGGTCC

[a] DuPw=wheat EST-SSR from DuPont
[b] Single-stranded nucleic acid

Table 4 Nucleotide motif, expected allele size, observed allele size range and the chromosomal location of the 42 SSRs used to genotype the durum accessions

Locus	Repeat	Allele size (bp)[a]	Allele size range	Chrom. location	Number of polymorphic alleles	Diversity index (DI)	Proportion of heterozygosity	Probability of identity (PI)
DuPw004[c]	(AC)13	199	193–203	4A	7	0.84	0.387	0.038
DuPw023[c]	(GCT)9	231	217–233	4B	5	0.71	0.375	0.126
DuPw038[c]	(GCC)9	193	184–201	1A	6	0.79	0.429	0.066
DuPw043[c]	(CAA)10	233	228–246	4B	6	0.80	0.536	0.062
DuPw108	(AT) 5(ATC)5	173	174–181	4A	2	0.35	0.016	0.476
DuPw115	(ACG)7	193	181–190	5B	3	0.53	0.234	0.264
DuPw123	(ACGT)5	199	180–191	NM[b]	1[d]	0.09	0.000	0.831
DuPw124[c]	(ACT)7	165	156–164	NM	3	0.63	0.155	0.175
DuPw135	(AGATG)5	246	175–247	NM	3	0.40	0.197	0.394
DuPw165	(AACAGC)5	168	169–172	NM	1[d]	0.29	0.000	0.544
DuPw167[c]	(AAGCAT)5	237	228–250	6A	4	0.78	0.286	0.071
DuPw168	(AATAG)12	167	149–166	NM	2	0.18	0.031	0.684
DuP173[f]	(AC)6 (AT)9	179	166–182	3D	2	0.09	0.000	0.830
DuPw205	(AAG)6	168	158–178	5B	5	0.63	0.180	0.162
DuPw207	(AAG)23	178	170–285	2B	1[d]	0.24	0.000	0.604
DuPw210	(AAG)6	227	228–232	2A	2	0.30	0.828	0.511
DuPw216	(AAG)9	186	164–200	6B	7	0.71	0.339	0.131
DuPw217[c]	(AAG)12	225	212–224	6B	5	0.84	0.483	0.046
DuPw227[c]	(AAG)7	193	167–203	3A	6	0.81	0.315	0.060
DuP238[f]	(AC)9	218	209–221	4D	5	0.73	0.241	0.114
DuPw254[c]	(AC)14	150	152–184	7A	8	0.76	0.190	0.083
DuPw398	(AGATG)5	189	177–198	7B	5	0.65	0.254	0.176
WMC017	$(CA)_{16}$	181	176–220	7A	7	0.63	0.017	0.204
WMC024	(GT)28	152	115–123	1A	3	0.61	0.164	0.197
WMC025	(GT)26	166	167–220	2B	3	0.48	0.091	0.278
WMC027	$(GT)_{26}TG(GT)_{2}$	389	383–399	NM	5	0.73	0.018	0.109
WMC047	$(CA)8GA(GA)_{5}$	141	138–144	4B	5	0.77	0.175	0.079
WMC076	$(GT)_{19}$	256	232–258	7B	3	0.40	0.111	0.384
WMC093	$(GT)_{24}$	149	232–258	NM	3	0.31	0.078	0.491
WMC106	$(CA)_{21}$	318	146–170	1A	8	0.84	0.383	0.045
WMC120	(CA)8 (GA)14 (GT)8	215	199–228	6B	2	0.20	0.000	0.652
GWM043[e]					3	0.03	0.000	0.498
GWM068					2	0.51	0.000	0.353
GWM099					5	0.51	0.016	0.184
GWM149					1[d]	0.62	0.000	0.940
GWM164					4	0.62	0.797	0.181
GWM169					14	0.96	0.452	0.009
GWM193					3	0.57	0.271	0.285
GWM304					5	0.71	0.091	0.198
GWM410					9	0.32	0.321	0.108
GWM526					8	0.63	0.127	0.153
GWM544					7	0.81	0.018	0.060

[a] According to ITMI parent Chinese spring
[b] NM=not mapped
[c] Selected as the best combination for genotyping
[d] Null allele detected
[e] Information on the GWM clones is given by Röder et al. 1998
[f] Detect polymorphic alleles in two genomes (P. Wolters, unpublished data)

dendrogram (Fig. 2). There does appear to be some clustering based on the geographical origin of the genotypes, with some evidence for the improved varieties also being grouped together. The dendrogram presented is constructed from 189 alleles detected with the three sources of SSRs. The coefficient of similarity ranged from 0.28 to 0.70, and varied when different sources of SSRs were used to generate the similarity matrix and estimates of phenetic distance. The similarity coefficient for the EST-SSR-derived similarity matrix (89 alleles) ranged from 0.37 to 0.72 compared to 0.26 to 0.79 for the combined data from the genomic SSRs (dendogram not shown).

The genotypes used in the present study include more than 60% of the total number of genotypes studied earlier by Sorrells et al. (1995) and Autrique et al. (1996) using RFLP and other markers. The SSR markers were more polymorphic than RFLPs and, although the same coefficient of similarity was applied, the genetic relationship between genotypes differed when SSR markers were deployed. For example, the RFLP (165 alleles) markers failed to discriminate between two landraces, Syria 3 and Syria 5, whereas the SSR markers placed these genotypes into two different groups with a similarity coefficient of 0.41. Additionally, in this study we used the SSR markers to estimate the proportion of heterozy-

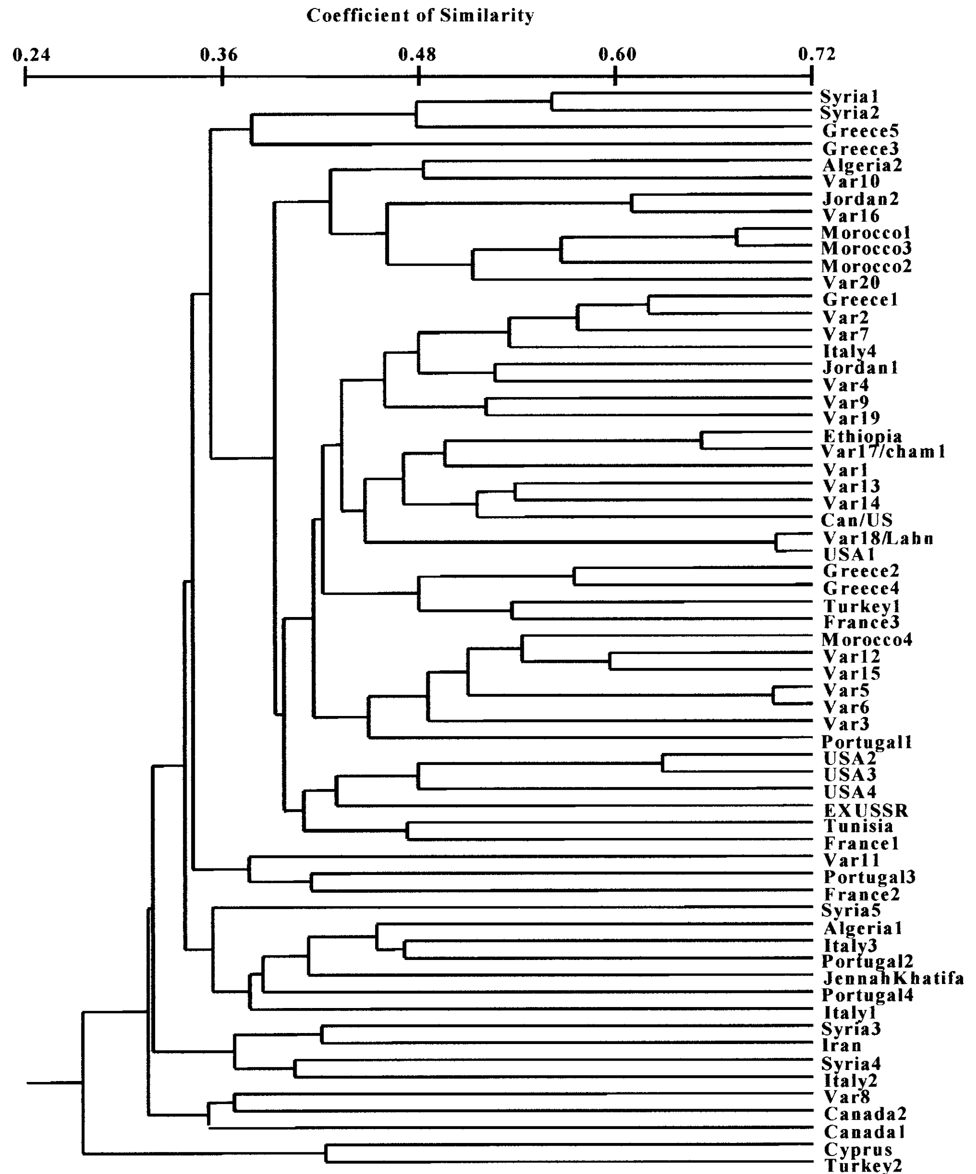


Fig. 2 Genetic relationship between 64 durum accessions, landraces and varieties, based on Jaccard's coefficient and UPGMA cluster analysis. The genotype name can be obtained from Table 1

gosity, which is an important parameter for landraces and breeding lines. For landraces such a measurement most-likely reflects residual heterogeneity.

Discussion

Genomic SSR markers have been used for discriminating between genotypes in bread wheat (Plaschke et al. 1995; Bryan et al. 1997; Prasad et al. 2000; Stachel et al. 2000) but the large genome size of wheat means that identifying robust and informative SSRs for fingerprinting is challenging. Our results show that with 20 SSR loci detecting approximately 100 alleles it was possible to discriminate all the durum genotypes studied. Furthermore, ten EST-SSRs provide an extremely low probability that any two genotypes will exhibit a chance match. These EST-derived SSRs generate high quality amplification products (Fig. 1), a high information content (diversity indices in excess of 0.76), and have been flourescentally labelled which allows detection on an automated DNA sequencer. This method of allele detection and sizing provides the potential advantage of multiplex analysis, automated data output and elimination of radioactivity. However, the SSRs described can also be detected reproducibly using isotopic labelling or silver staining.

Two other studies have examined the informativeness of SSRs derived from ESTs in rice (Cho et al. 2000) and grape (Scott et al. 2000). These two studies concur with the present investigation in revealing that the overall level of informativeness of genomic-derived SSRs is higher than that observed with EST-derived SSRs. Cho et al. (2000) found that GC-rich tri-nucleotide motifs located in exons or open reading frames showed considerably lower levels of polymorphism as compared to microsatellites found in the 3′ or 5′ untranslated regions

or introns. Scott et al. (2000) were also able to show that SSRs within coding regions were polymorphic and transferable between species. Similarly, the sample of wheat EST-derived SSRs evaluated in this study identified informative SSRs. Perhaps, more importantly, mutationally stable SSRs that are less polymorphic and derived from ESTs could complement highly variable genomic SSRs to reconstruct past evolutionary events and identify regions of plant genomes identical by descent. Differences in mutation rate between di, tri, and tetra-nucleotide repeat motifs have been observed in humans (Chakraborty et al. 1997). In the present study several higher-order repeat motifs were identified in the sample of EST-derived wheat SSRs (Table 4). The tri, tetra and penta-nucleotide repeats are likely to exhibit varying mutational properties compared to di-nucleotide repeats and also offer considerable advantages for allele sizing.

The identification and exploitation of EST-derived SSRs in wheat also represents a shift in emphasis from anonymous markers to reveal variation in transcribed regions of wheat genomes. Expansion and contraction of SSR repeats in genes of known function can now be related to phenotypic variation or, more desirably, biological function. Starch composition in the rice kernel has been shown to be associated with expansion of a $(CT)_n$ repeat in the 5′ UTR of the waxy gene (Ayers et al. 1997). The more-widespread application of EST-derived SSRs to the evaluation of germplasm collections will provide an opportunity to examine functional diversity in phenotypically well -characterised accessions. The power of this approach will be augmented when meiotic map-location and putative quantatitive trait data are combined with the allelic variability detected in various gene pools.

Ultimately, bi-allelic markers in the form of single nucleotide polymorphisms (SNPs) detected in candidate genes will provide the basis for linkage disequilibrium or association mapping. The exploitation of historical recombination events identified in germplasm collections, coupled with SNP-based haplotype detection, will drive future map-based gene-discovery programmes. Various methods for SNP detection have been demonstrated (Kuklin et al. 1997; Chen et al. 1998; Cho et al. 1999; Erdos et al. 1999), but the routine detection of SNPs in polyploid species such as wheat remains challenging whereas the EST-derived SSRs reported in this study provides a resource which can be immediately applied to many aspects of wheat genetics. As large numbers of publicly available ESTs are generated for other plant species, it is anticipated that EST-derived SSRs will represent an important source of informative and robust polymorphic markers.

Acknowledgements Imad Eujayl was financially supported by BMZ (Grant No. 98.7860.4-001.00) and the Arab Fund (AFSED - Kuwait). ICARDA acknowledges the substantial input of E. I. DuPont De Nemours in the capital and consumables to make this work possible. We thank Dr. M. Röder (IPK/Gatersleben-Germany) and the Agrogene Wheat Microsatellite Consortium for providing SSR primer information. The experiments conducted in this study comply with the current laws and regulations of the USA.

References

Adams MD, Dubnick M, Kerlavage AR, Moreno R, Kelley JM, Utterback TR, Nagle JW, Fields C, Venter JC (1992) Sequence identification of 2375 human brain genes. Nature 355:632–634

Ajioka JW (1998) *Toxoplasma gondii*: ESTs and gene discovery. Int J Parasitol 28:1025–1031

Altschul SF, Gish W, Miller W, Myers EW, Lipman D (1990) Basic local alignment search tool. J Mol Biol 215:403-410

Autrique E, Nachit MM, Monneveux P, Tanksley SD, Sorrells ME (1996) Genetic diversity in durum wheat based on RFLPs, morphophysiological traits and coefficient of parentage. Crop Sci 36:735-742

Ayers NM, McClung AM, Larkin PD, Bligh HFJ, Jones CA, Park WD (1997) Microsatellites and a single nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germplasm. Theor Appl Genet 94:773–781

Blanco A, Bellomo MP, Cenci A, De Giovanni C, D'Ovidio R, Iacono E, Laddomada B, Simeone R, Tanzarella OA, Porceddu E (1998) A genetic linkage map of durum wheat. Theor Appl Genet 97:721–728

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Munch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455–457

Briggs SP (1998) Plant genomics: More food for thought. Proc Nat Acad Sci USA 95:1986–1988

Bryan GJ, Collins AJ, Stephenson P, Smith JB, Orry A, Gale MD (1997) Isolation and characterisation of microsatellites from hexaploid bread wheat. Theor Appl Genet 94:557–563

Chakraborty R, Kimmel M, Strivers DN, Davison LJ, Deka R (1997) Relative mutation rates at di- tri- and tetra-nucleotide microsatellite loci. Proc Nat Acad Sci USA 94:1041–1046

Chen XN, Livak KJ, Kwok PY (1998) A homogeneous, ligase-mediated DNA diagnostic test. Genome Rese 8:549–556

Cho RJ, Mindrinos M, Richards DR, Sapolsky RJ, Anderson M, Drenkard E, Dewdney L, Reuber TL, Stammers M, Federspiel N, Theologis A, Yang WH, Hubbel E, Au M, Chung EY, Lashkari D, Lemieux B, Dean C, Lipshutz RJ, Ausubel FM, Davis RW, Oefner PJ (1999) Genome-wide mapping with biallelic markers in *Arabidopsis thaliana.* Nature Genet 23:203–207

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Parl WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza satica* L.). Theor Appl Genet 100:713–722

Delseney M, Cooke R, Raynal M, Grellet F (1997) The *Arabidopsis thaliana* cDNA sequencing project. FEBS Lett 405: 129–132

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational process of simple-sequence repeat loci in human populations. Proc Nat Acad Sci USA 91:3166–3170

Erdos MR, Braun A, Mihike K, Silander K, Ghosh S, Collins FS (1999) MALDI-TOF mass spectrometry for high throughput SNP detection: application to type-2 diabetes. Am J Human Genet 65:1381

Korzun V, Röder MS, Wendehake K, Pasqualone A, Lotti C, Ganal MW, Blanco A (1999) Integration of dinucleotide microsatellites from hexaploid bread wheat into a genetic linkage map of durum wheat. Theor Appl Genet 98:1202–1207

Kuklin A, Munson K, Gjerde D, Haefele R, Taylor P (1997) Detection of single-nucleotide polymorphisms with the WAVE (TM) DNA fragment-analysis system. Genet Testing 1:201–206

Paetkau D, Calvert W, Stirling I, Strobeck C (1995) Microsatellite analysis of population structure in Canadian polar bears. Mol Ecol 4:347–354

Plaschke P, Ganal MW, Röder MS (1995) Detection of genetic diversity in closely related bread wheat using microsatellite markers. Theor Appl Genet 91:1001–1007

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 7:215–222

Prasad M, Varshney RK, Roy JK, Balyan HS, Gupta PK (2000) The use of microsatellites for detecting DNA polymorphism, genotype identification and genetic diversity in wheat. Theor Appl Genet 100:584–592

Röder MS, Plaschke J, König SU, Börner A, Sorrells ME, Tanksley SD, Ganal MW (1995) Abundance, variability and chromosomal location of microsatellites in wheat. Mol Gen Genet 246:327–333

Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier M, Leroy P, Ganal M (1998) A microsatellite map of wheat. Genetics 149:2007–2023

Rohlf, FJ (1990): NTSYS-pc manual. Exeter Software, Setauket, New York

Scott KD, Eggler P, Seaton G, Rosetto M, Ablett EM, Lee LS, Henry RJ (2000) Analysis of SSRs derived from grape ESTs. Theor Appl Genet 100:723–726

Shure M, Wessler S, Fedoroff N (1983) Molecular identification and isolation of the Waxy locus in maize. Cell 35:225–233

Sorrells ME, Barbosa J, Nachit MM, Ketata H, Autrique E (1995) Relationships among 81 durum genotypes based on RFLPs, gliadins, parentage and quality traits. Options Mediterraneenes, Serie A, Seminaires Mediterraneens, 22:249–262

Stachel M, Lelley T, Grausgruber H, Vollmann J (2000) Application of microsatellites in wheat (*Triticum aestivum L.)* for studying genetic differentiation caused by selection for adaptation. Theor Appl Genet 100:242–248

Takahashi N, Ko MS (1994) Toward a whole cDNA catalogue: construction of an equalised cDNA library from mouse embryos. Genomics 23:202–210

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res 17:6463–6471

Tautz D, Renz M (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res 12:4127–4138

Yamamito K, Sasaki T (1997) Large-scale EST sequencing in rice. Plant Mol Biol 35:135–144

Vassarotti A, Goffeau A (1992) Sequencing the yeast genome: the European effort. Trends Biotechnol 10:15–18

Waterston R, Martin C, Craxton M, Huynh C, Coulson A, Hillier L, Durbin R, Green P, Shownkeen R, Halloran N (1992) A survey of expressed genes in *Caenorhabditis elegans*. Nature Genet 1:114–123

Weir BS (1990) Genetic data analysis. Sinauer Associates, Sunderland, Massachusetts