Theor Appl Genet (2003) 106:411–422
DOI 10.1007/s00122-002-1031-0

T. Thiel · W. Michalek · R.K. Varshney · A. Graner

Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.)

Received: 9 April 2002 / Accepted: 21 May 2002 / Published online: 14 September 2002

Abstract A software tool was developed for the identification of simple sequence repeats (SSRs) in a barley (*Hordeum vulgare* L.) EST (expressed sequence tag) database comprising 24,595 sequences. In total, 1,856 SSR-containing sequences were identified. Trimeric SSR repeat motifs appeared to be the most abundant type. A subset of 311 primer pairs flanking SSR loci have been used for screening polymorphisms among six barley cultivars, being parents of three mapping populations. As a result, 76 EST-derived SSR-markers were integrated into a barley genetic consensus map. A correlation between polymorphism and the number of repeats was observed for SSRs built of dimeric up to tetrameric units. 3′-ESTs yielded a higher portion of polymorphic SSRs (64%) than 5′-ESTs did. The estimated PIC (polymorphic information content) value was 0.45 ± 0.03. Approximately 80% of the SSR-markers amplified DNA fragments in *Hordeum bulbosum*, followed by rye, wheat (both about 60%) and rice (40%). A subset of 38 EST-derived SSR-markers comprising 114 alleles were used to investigate genetic diversity among 54 barley cultivars. In accordance with a previous, RFLP-based, study, spring and winter cultivars, as well as two- and six-rowed barleys, formed separate clades upon PCoA analysis. The results show that: (1) with the software tool developed, EST databases can be efficiently exploited for the development of cDNA-SSRs, (2) EST-derived SSRs are significantly less polymorphic than those derived from genomic regions, (3) a considerable portion of the developed SSRs can be transferred to related species, and (4) compared to RFLP-markers, cDNA-SSRs yield similar patterns of genetic diversity.

Keywords Simple sequence repeats (SSRs) · Polymorphism information content (PIC) · Genetic mapping · Diversity analysis

Communicated by G. Wenzel

T. Thiel · W. Michalek · R.K. Varshney · A. Graner (✉)
Institute of Plant Genetics and Crop Plant Research (IPK),
Corrensstrasse 3, D-06466 Gatersleben, Germany
e-mail: graner@ipk-gatersleben.de
Tel.: +49-39482 5521, Fax: +49-39482 5155

Present address:
W. Michalek, PLANTA GmbH, Grimsehlstrasse 31,
D-37555 Einbeck, Germany

Introduction

Microsatellites, or simple sequence repeats (SSRs), are stretches of DNA consisting of tandemly repeated short units of 1–6 base pairs in length. The uniqueness and the value of microsatellites arise from their multiallelic nature, codominant inheritance, relative abundance, extensive genome coverage and simple detection by PCR using two unique primers, that flank the microsatellite and hence define the microsatellite locus (Powell et al. 1996).

The standard method for developing SSR-markers involves the creation of a small-insert genomic library, the subsequent hybridization with tandemly repeated oligonucleotides and the sequencing of candidate clones thus making the process time consuming and labor-intensive. In barley, several hundreds of microsatellite markers have been developed using this strategy (Liu et al. 1996; Struss and Plieske 1998; Ramsay et al. 2000).

An alternative strategy arises from increasing information available in DNA sequence databases. SSRs can be searched in these databases, thus shortening time and costs required for their development. However, the major drawback was the availability of sufficient sequence information. For that reason only a few SSR-markers have been developed in this manner so far (Saghai Maroof et al. 1994; Becker and Heun 1995; Pillen et al. 2000). Due to the rapid increase of sequence information the generation of EST-derived microsatellite markers becomes an attractive alternative to complement-existing SSR collections. In this context, the use of EST or cDNA-based SSRs has been reported for several species including grape (Scott et al. 2000), sugarcane (Cordeiro et al.

Table 1 Pedigree information of 25 European winter and 29 European spring barley cultivars used in SSR analysis. Reference: Baumer and Cais 2000; Abbreviations: 'F.Berg' = 'Friedrichswerther Berg'; 'Vogels. Gold' = 'Vogelsanger Gold'; IPK = Institut für Pflanzengenetik und Kulturpflanzenforschung, Gatersleben; BAZ = Bundesanstalt für Züchtungsforschung an Kulturpflanzen, Braunschweig

No.	Cultivar	Pedigree	Accession
Winter type, six-rowed			
1	Adonia	Triesdorfer Stamm × Dea	IPK – H HOR 3064/75
2	Brunhild	Barbo × Banteng	BAZ – 64024
3	Catinka	(Dura × Vogels. Gold) × Mirra	BAZ – 65272
4	Corona	[(Dura × Dea) × Perga] × Vogels. Gold	IPK – H HOR 10216/85
5	Dea	[(Ragusa × Peragis12) × (Heils Franken × Frw.Berg)] × [(Ragusa × Mahnd.Viktoria) × (Bolivia × Ragusa)]	IPK – H HOR 3245/76
6	Dura	(Friedrichswerther Berg × Ragusa) × Doria	BAZ – 4887
7	Franka	[(Vogels. Gold × Senta) × (Dura × Dea)] × Vogels. Gold	IPK – H HOR 10217/85
8	F. Berg	Groninger × Eckendorfer Mammuth	IPK – H HOR 2272/93
9	Ginso	(Dura × Senta) × Vogels.Gold	IPK – H HOR 9896/93
10	Mammut	Vogels.Gold × (Mädru × Weissenhaus 382/49)	IPK – H HOR 10598/98
11	Vogels. Gold	[Isaria × H204(*Hord.sp.nigrum*)] × WG 5	IPK – H HOR 4432/90
Winter type, two-rowed			
12	Cosima	Hydra × LBP 4139	BAZ – 63434
13	Danilo	Marko × FDE 257-21-5-1	IPK – H HOR 10476
14	Filia	Igri × LBP 5098	BAZ – 58622
15	Igri	Malta × LBP 1427 × Ingrid	IPK – H HOR 10596/88
16	Kaskade	Carsten 2zlg. × Union × Malta × Emir	IPK – H HOR 10029/98
17	Malta	Carsten 1565 × Strengs Aurea (SG) × Dea × Herfordia	IPK – H HOR 4440/90
18	Palazia	LBP1927(Malta × LBP 1423) × LBP 9572(Iduna × LBP 9375)	BAZ – 62935
19	Posaune	Sonja × (Hydra × LBP 3533)	BAZ – 62934
20	Sonate	Sonja × Stamm	BAZ – 63429
21	Sonja	Tria × Malta	IPK – H HOR 10595/97
22	Timura	Igri × LBP 1911	BAZ – 58619
23	Tria	(LBP 348/342 × Firlbecks 4zlg.) × Carsten 1565	BAZ – 35701
24	Trixi	Ber 685 × LBP 2437 × LBP 4072	BAZ
25	Viola	LBP 4622 × (Malta × Sonja)	BAZ – 63430
Spring type, two-rowed			
26	Alexis	Br.1622 × Trumpf	IPK – A HOR 11391/96
27	Aramir	Volla × Emir	IPK – A HOR 4965/82
28	Arena	Amethyst × Aufhammer 39/68	IPK – A HOR 10396/86
29	Barke	Libelle × Alexis	IPK – A HOR 13170
30	Berolina	Medina × Trumpf	IPK – A HOR 10397/86
31	Carina	(Union × W 16 WV) × Volla	IPK – A HOR 4947/87
32	Cheri	Trumpf × (Medusa × Diamant)	BAZ – 64016
33	Defra	Gerlinde × Karat	IPK – A HOR 10288/85
34	Dorett	Amethyst × Aufhammer 39/68	IPK – A HOR 10469/86
35	Emir	Delta × Agio × Kenia 3 × Arabische	IPK – A HOR 3548/96
36	Gimpel	Proctor × Carlsberg II × Heine 4808 × Stamm	IPK – A HOR 9465/82
37	Golf	(Armelle × Lud) × Luke	IPK – A HOR 10173/85
38	Grit	(55474/67 × 46459/68) × 480/68	IPK – A HOR 8899/84
39	Haisa I	Heines Hanna × Ackermann Isaria	BAZ – 5506
40	Hockey	Claret × Goblin	IPK – A HOR 10174/85
41	Isaria Nova	Heils Franken × (Danubia × Isaria)	IPK – A HOR 2533/83
42	Koral	Hanna × [(Carlsb.II × Union) × Alsa] × Hanacky × J25	IPK – A HOR 8837/85
43	Libelle	(Br.1747 × Rupee) × Br.1622	IPK – A HOR 11392/98
44	Lilo	Abed 63944 × Herta8 × 191 × Ingrid × Minerva × Aramir	BAZ – 5603
45	Ragusa	Selection from the Dalmatian landrace	IPK – A HOR 2267/98
46	Teo	Claret × Kym	BAZ – 65262
47	Toga	Trumpf × Welam	BAZ – 62039
48	Trumpf	Diamant × 14029/64/6 [(Alsa × S3170/Abyss) × 11719/59)] × Union	IPK – A HOR 4967/96
49	Ultra	Aramir × Carina	IPK – A HOR 10394/86
50	Union	(Weihenst.MR II × Donaria) × Firl. 621	BAZ – 5472
51	Ursel	Aramir × Trumpf	IPK – A HOR 10468/86
52	Villa	Volla × Haisa I × Wisa	IPK – A HOR 11352/97
53	Volla	Wisa × Haisa I	BAZ – 5466
54	Wisa	(Weihenst.MR I × Breun IN 2511) × Isaria	BAZ – 5463

2001), durum wheat (Eujayl et al. 2002) and rye (Hackauf and Wehling 2002).

Keeping in view the above, the present study was undertaken with the objectives: (1) to develop a software-tool to be used in mining of EST-databases for the presence of SSR-motifs, (2) to develop and characterize a collection of EST-derived SSR-markers for barley in terms of frequency, information content, genomic distribution and transferability to related species, and (3) to assess their potential for diversity analysis in a reference set of European barley cultivars.

Materials and Methods

Plant material

To screen for polymorphic SSRs, the barley cultivars 'Igri', 'Franka', 'Steptoe', 'Morex', Oregon Wolfe Barley 'OWBrec' and 'OWBdom' were used, which are parents of three doubled-haploid (DH) mapping populations.

A total of five wild barley (*Hordeum bulbosum*) accessions ('A17', 'HB2032', 'Cb2920/4', 'Cb2929/1' and 'GRA951.80'), two rye (*Secale cereale* L.) accessions ('P87' and 'P105') and one accession each of wheat (*Triticum aestivum* L. cv 'Chinese Spring') and rice (*Oryza sativa* L.) were used for the examination of the transferability of barley EST-derived SSR-markers. For diversity analysis a total of 54 barley cultivars (as described in Table 1) and one wild barley (*Hordeum spontaneum*) accession obtained from the IPK genebank (accession no. HOR 11508) were used. Genomic DNA from the above plants was isolated from young leaves using the CTAB method described by Saghai-Maroof et al. (1984). For DNA extraction, leaves from about ten individual seedlings per accession or DH-line were pooled.

In silico data mining for SSR-markers

A total of 24,595 barley EST sequences corresponding to approximately 12.8 Mb were screened for microsatellites. ESTs comprising both 5′- and 3′-sequences were developed from cDNA libraries representing 16 different tissues or developmental stages, with the vast majority of sequences being derived from the cultivar 'Barke'. A detailed description of a subset of the ESTs used for this study is given by Michalek et al. (2002). In a preliminary step, polyA and polyT stretches which correspond to polyA-tails in eucaryotic mRNA were removed with the help of a PERL5 script until no stretch of $(T)_5$ or $(A)_5$ was present in a window of 50 bp on the 5′- or 3′-end, respectively. EST sequences of less than 100 bp were rejected and sequences larger than 700 bp were clipped at their 3′ side to preclude the inclusion of low quality sequences in the further analysis. The identification and localization of microsatellites was accomplished by another PERL5 script (named as *MI*cro*SA*tellite MISA) able to identify both perfect microsatellites as well as compound microsatellites which are interrupted by a certain number of bases.

In this study, microsatellites were considered to contain motifs that are between one and six nucleotides in size. Thereby the minimum length criteria was defined as being ten repeat units for mononucleotides, six repeat units for dinucleotides and five repeat units for all higher-order repeats.

Once SSR-containing ESTs have been identified, flanking primers were designed using PRIMER3 (Rozen and Skaletsky 1998) run in a batch modus with the help of PERL5 interface modules. To force the selection of flanking primers, the 'target' option was used representing the position of the respective microsatellite enlarged by three positions at each side. All other options were left at their default value. Thus, primers were designed defining loci ranging from 100 to 300 bp in length. For the primers, the value of the optimal melting temperature was set to 60 °C, and the optimal size to 20 bases. In order to eliminate redundancy, a cluster analysis was performed using stackPACK2.1 software run under Solaris (Miller et al. 1999).

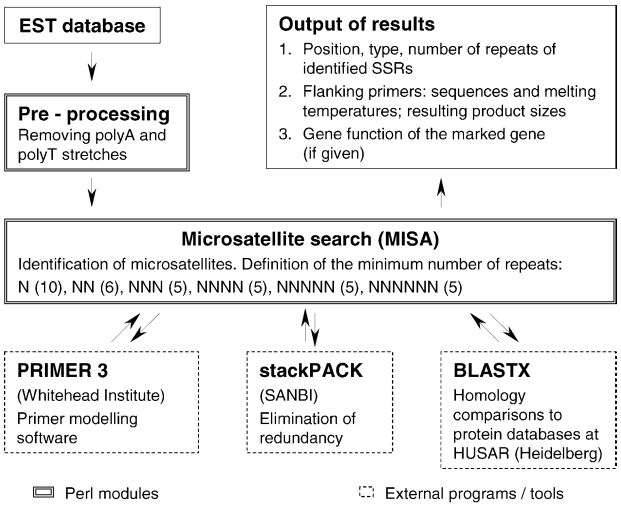

Fig. 1 Data mining for the development of EST-derived microsatellite markers. In conjunction with a set of additional software-programs (Primer 3, stackPACK, BlastX), the Microsatellite search module (MISA) identifies SSR-containing ESTs from an input database together with primer sequences for a non-redundant set of SSRs and data about putative functions

In order to obtain an idea about putative functions of SSR-containing genes, these sequences were compared to the SWISSPIR-PLUS protein database (as of July 2001) at the DKFZ Heidelberg using the BlastX2 program (Altschul et al. 1997) assuming an e-value < 1E-20 as a significant homology (for details see http://genome.dkfz-heidelberg.de/). BlastN searches were conducted against wheat and rice ESTs to address the question which proportion of barley SSR loci have homologies to ESTs from related species. Publicly available EST sequences were acquired by the Sequence Retrieval System (SRS6, http://srs.ebi.ac.uk/) from the EMBL database (query: organism = '*Oryza sativa*'/'*Triticum aestivum*', and division = 'EST') comprising 81,328 and 67,470 entries, respectively (November 2001). Additionally, loci were blasted against publicly available marker sequences for rice from the RGP and Gramene using TIGR Rice BLAST (http://www.tigr.org/).

A flow chart summarizing the individual steps of the data-mining procedure is given in Fig. 1. The software allows an individual adjustment of the criteria described above for the identification of microsatellite motifs and the positioning of flanking primers, thus complying with the individual needs of a particular experiment. The final output summarizes all important information about the position and type of microsatellites identified, as well as experimentally important outputs of Primer3, e.g. the primer sequences or the expected size of the PCR product. Additionally, frequencies of the individual SSR motifs of the total set are estimated taking into regard the complementary sequence, for instance the motifs AC, CA, TG and GT that belong to the same class. The Microsatellite search module (MISA) is available under http://pgrc.ipk-gatersleben.de/misa/.

PCR conditions and separation of microsatellites

PCR was carried out in 10-µl reactions consisting of a 1 × PCR buffer (Qiagen, Hilden) including 1.5 µM $MgCl_2$, 200 µM dNTPs, 250 nM of each primer, 0.25 U of *Taq* DNA polymerase (Qiagen, Hilden), and 20 ng of genomic DNA. All fragments were amplified using the following touch-down PCR profile: an initial dena-

turing step of 3 min at 94 °C was followed by 45 cycles with denaturation at 94 °C for 30 s and extension at 72 °C for 30 s, respectively. The annealing temperature was decreased in 0.5 °C increments from 60 °C in the first cycle to 55 °C after the 10th cycle and was then kept constant for the remaining 35 cycles (always 30 s). After 45 cycles a final extension step was performed at 72 °C for 5 min. PCR amplifications were carried out using a GeneAmp 9700 cycler (Perkin Elmer, Weiterstadt).

PCR fragments were separated on denaturing polyacrylamide gels consisting of 10% polyacrylamide (AA:BIS = 19:1) and 7 M urea in 0.5 × TBE buffer. To this end PCR reactions were mixed with equal volumes of loading buffer (formamide containing 0.8 mM EDTA and traces of bromophenol blue and xylene cyanol), denatured at 95 °C for 5 min and snap cooled on ice. Afterwards, samples were loaded on pre-heated Sequi-Gen GT Sequencing Cells (Bio Rad, Munich), which were run at 1,800 V for 2.5 up to 4 h, depending on the fragment sizes to be separated.

After the run, the fragments were visualized by silver staining. For this, the gel was fixed for 5 min in fixative (10% ethanol, 5% acetic acid), rinsed in deionized water, stained for 15 min in 0.3% (w/v) silver nitrate, rinsed again in deionized water, and developed for approximately 15 min until the bands became visible in 1.5 g/l of NaOH, 4 ml/l of 37% formaldehyde and 85 µg/l of NaBH. Scoring was done by visual inspection.

Linkage mapping

Polymorphic SSRs were genetically mapped using three populations. The first population was constructed from the cross 'Igri' × 'Franka' comprising 74 doubled-haploid (DH) lines (I/F, Graner et al. 1991). The second consisted of the 150 DH lines of a cross between the cultivars 'Steptoe' and 'Morex' (S/M, Kleinhofs et al. 1993). The 94 DHs of the third mapping population were derived from a cross of two Oregon Wolfe Barley (OWB) marker stocks 'Rec(essive)' and 'Dom(inant)', subsequently designated 'OWBrec' and 'OWBdom' (Costa et al. 2001).

Initially, SSRs were integrated in one of the above-mentioned maps using the MAPMAKER Macintosh V2.0 program (Lander et al. 1987) with the help of the 'try' command, followed by rechecking the marker order with the 'ripple' command. In order to construct a consensus linkage map out of the three individual maps, the program JoinMap (Stam 1993) was used (LOD = 0.001; rec = 0.499) employing the 'fixed order' option in a similar way as described in Qi et al. (1996). To this end, mainly anchor markers were chosen being mapped in at least two populations. With regard to estimating the localization of the centromeres, appropriate markers were included from the map of Künzel et al. (2000). The Kosambi function was chosen to calculate map distances (Kosambi 1944). The order of markers on the consensus map was finally re-checked with the individual maps. In a few cases, observed segregations of previously co-segregating markers (maximal 1 cM) were corrected manually. Marker loci are designated as GBM1001–GBM1076 (*G*atersleben *B*arley *M*icrosatellite).

Statistical analysis

The polymorphic information content (PIC) is a tool to measure the informativeness of a given DNA marker. According to Weber (1990) and Anderson et al. (1993) the PIC-value was calculated as follows:

$$\mathrm{PIC} = 1 - \sum_{i=1}^{k} P_i^2,$$

where k is the total number of alleles detected for a microsatellite and P_i the frequency of the *i*th allele in the set of six barley cultivars investigated.

For diversity analysis, fragments were scored as present (1) or absent (0) for each of the 55 barley cultivars and accessions. Genetic similarities (GS) between each pair were calculated according to Nei and Li (1979) by using the following formula:

$$\mathrm{GS}_{ij} = \frac{2a}{2a+b+c}$$

where *a* is the number of shared fragments, *b* and *c* the number of alleles being present either in line *i* or line *j*, respectively.

Based on estimated genetic dissimilarities a Principal Coordinates Analysis (PCoA) was performed using the NTSYSpc software package (version 2.02, Rohlf 1997) and the consecutive commands 'Dcenter', 'Eigen' and the '3D plot' in order to generate the three-dimensional PCoA plot.

Results

Frequency and distribution of microsatellites

The search for microsatellites in 24,595 barley EST sequences representing approximately 12.8 Mb revealed 2,019 microsatellites in 1,856 ESTs; 136 ESTs contained more than one SSR. This corresponds to an average distance between SSRs of approximately 6.3 kb or one SSR-containing EST every 13.3 ESTs. The average number of repeats across all motifs analyzed was 7.3 ± 1.1. With a 17.1-bp size, microsatellites in EST sequences are rather small (84% of all SSRs were smaller than 21 bp); 2.1% of the SSRs were of a compound type. Eleven cases were found where two microsatellites were immediately adjacent to each other; 31 ESTs contained two adjacent repeats at a distance of <10 bp to each other.

For reasons of a probable overestimation of specific SSR types and in terms of marker development (see later), a redundancy analysis was performed on SSR-containing EST sequences. However, when comparing the relative frequencies of individual SSR motifs between the non-redundant SSRs with the redundant SSRs, no significant differences were observed.

The occurrence of the individual SSR motifs among the non-redundant set of 832 SSRs which were found in 752 EST sequences is summarized in Table 2. Obviously, the proportion of SSR unit sizes was not evenly distributed: 154 (19%) were mononucleotide, 153 (18%) dinucleotide, 470 (56%) trinucleotide, 46 (6%) tetranucleotide, 5 penta- and 4 hexa-nucleotide microsatellites. Regarding dimeric SSRs, the motifs AG (51%) and AC (33%) were by far the most-common ones, whereas AT and CG microsatellites were present only at low abundance (10% and 6%, respectively). Among trimeric microsatellites, CCG (36%), AGG (15%) and AGC (14%) were the most-common motifs. The most-frequent tetrameric microsatellite motifs were ACGT and ACAT.

SSR-marker development

Out of 1,856 microsatellite containing ESTs, primer pairs could be designed for 1,453 of them (78%). The remaining sequences contained either too little DNA se-

Table 2 Occurrence of non-redundant SSRs in a set of 24,595 barley ESTs

SSR Motif	Number of repeats												Total
	5	6	7	8	9	10	11	12	13	14	15	>15	
A/T	–	–	–	–	–	60	19	9	9	3		39	139
C/G	–	–	–	–	–	4	2	1	2	1	2	3	15
AC/GT	–	16	15	6	4	3	3	1			1	1	50
AG/CT	–	26	18	12	7	5	2		3		1	5	79
AT	–	9	4		2								15
CG	–	7		1			1						9
AAC/GTT	5	3			1		1						10
AAG/CTT	19	8	5	1	1								34
AAT/ATT	3											1	4
ACC/GGT	17	6	3	2									28
ACG/CTG	26	14	12	1									53
ACT/ATG	7	3	1	1	2								14
AGC/CGT	42	16	3	3	3								67
AGG/CCT	45	19	5		2								71
AGT/ATC	12	3	2	1									18
CCG/CGG	107	27	19	14	4								171
ACGT/ATGC	7												7
ACAT/ATGT	3		2						1				6
AACC/GGTT	4												4
AGGG/CCCT	2	2											4
AGGT/ATCC	2	1		1									4
AAGG/CCTT	1	1	1										3
Other NNNN	17	1											18
N	–	–	–	–	–	64	21	10	11	4	2	42	154
NN	–	58	37	19	13	8	6	1	3		2	6	153
NNN	283	99	50	23	13		1					1	470
NNNN	36	5	3	1					1				46
NNNNN	4	1											5
NNNNNN	4												4
												Total	832

quence flanking the microsatellite or the sequences were inappropriate for primer modelling. Redundancy analysis of those 1,453 sequences revealed 945 ESTs belonging to 244 clusters and 508 ESTs being singletons. Hence, a total of 752 unique sequences (3.1% of all EST sequences investigated) represented potential candidates for SSR-marker development.

Three hundred and eleven primer pairs flanking di-, tri-, or tetra-meric SSRs were assayed to detect polymorphisms in the six barley cultivars 'Igri', 'Franka', 'Steptoe', 'Morex', 'OWBrec' and 'OWBdom'. The results are summarized in Fig. 2. In 200 (64%) cases, fragments could be amplified during PCR with genomic DNA, while for 111 primer pairs PCR completely failed or led to the weak amplification of background signals leading to the exclusion from further analysis. Null alleles were detected with three primer pairs. PCR products of expected size were obtained in 155 cases. Of these approximately 17% of the primer pairs amplified additional, but in most cases weaker, fragments. In 45 cases amplicons were obtained which were of larger size than expected from the EST sequence, probably due to the presence of introns. The amplification of introns may cause problems, since fragments above 500 bp (as was the case for 20 primer pairs) cannot be scored accurately for small differences in fragment size. Additionally, it can be assumed that in several cases the observed polymorphism is caused by a size polymorphism within the intron, which may overshadow a putative polymorphism of the microsatellite. Ninety eight (54%) primer pairs were monomorphic between the six barley cultivars examined. Six primer pairs amplified fragments that appeared as two alleles, but were not polymorphic between parents

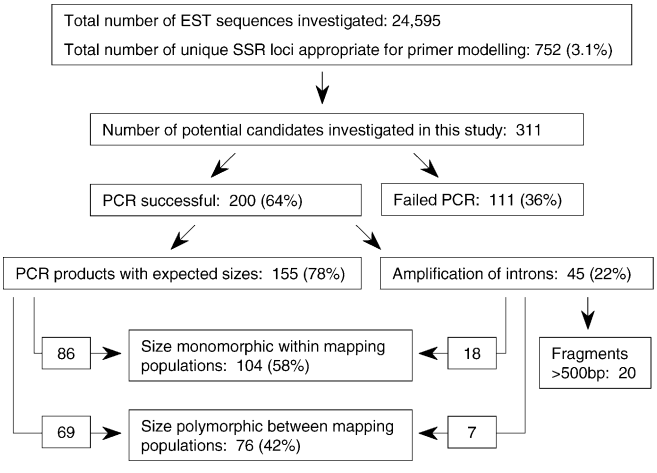

Fig. 2 Schematic overview summarizing the outcome of the individual steps of SSR-marker development

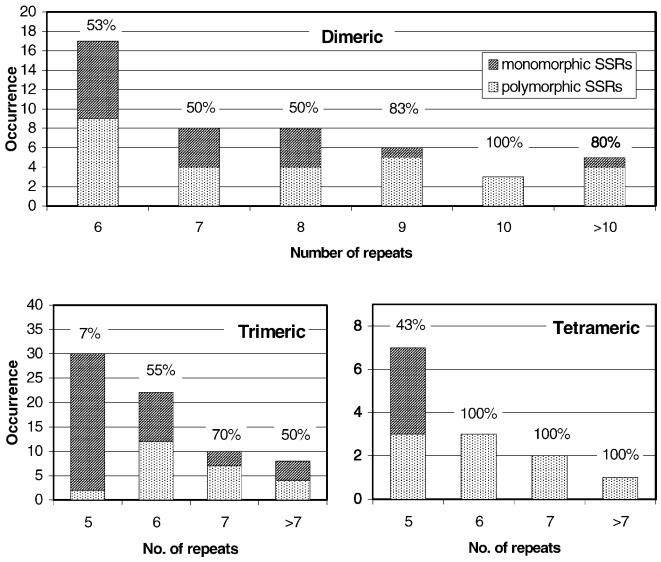

Fig. 3 Relationship between polymorphism and size of microsatellite loci. Only fragments of predicted size containing a single microsatellite were considered

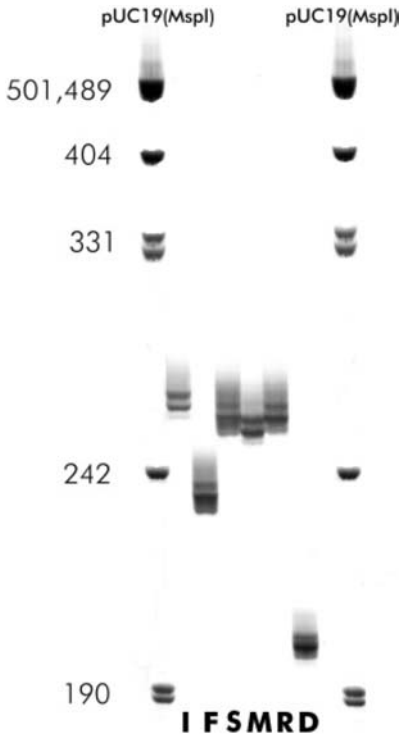

Fig. 4 EST-derived SSR-marker GBM1015. This highly polymorphic marker reveals a total of five alleles in six barley cultivars. I='Igri', F='Franka', S='Steptoe', M='Morex', R='OWBrec', D='OWBdom'. DNA size-markers (in bp) are shown on both sides of the gel

Table 3 Exploiting 3′ and 5′ ESTs for the development of polymorphic SSRs. ESTs containing different SSR types, as well as markers that amplify introns when performing PCR with genomic DNA, were excluded. Loci with multiple SSRs of the same type were counted once

SSR type	Dimeric	Trimeric	Tetrameric	Total
in 3′ ESTs	27 (77%)	13 (42%)	9 (90%)	49 (64%)
in 5′ ESTs	8 (23%)	18 (58%)	1 (10%)	27 (36%)
Total	35	31	10	76

of the mapping populations. The remaining 76 primer pairs yielded size polymorphic fragments (including seven cases where introns were amplified) between the parents of at least one of the mapping populations, with an average of 2.5 ± 0.2 alleles per locus in the six accessions analyzed.

SSR polymorphism

Eighty percent of the markers showed polymorphisms between 'OWBrec' and 'OWBdom', followed by S/M (53%) and I/F (21%). In total, 54% of the polymorphic markers were mappable in only one population, 38% in two and 8% of the markers were polymorphic in all three mapping populations. Table 3 summarizes SSR frequencies in 5′- and 3′-ESTs. Over all, 64% of the polymorphic SSRs were present in 3′-sequences. Based on the size of individual repeat units, 3′-ESTs were a more-rewarding resource for marker development of dimeric (77%) and tetrameric (90%) SSRs, while for trimeric repeats 5′-sequences seem to be slightly better (58%). Out of those markers, which contain more than one SSR (the PCR product is of predicted size), ten were polymorphic and one marker was monomorphic. The relationship between polymorphism and the number of repeats for investigated SSR types is given in Fig. 3. Among dimeric microsatellites the proportion of polymorphic microsatellites increases significantly from SSRs built of 6–8 repeats (approximately 50%) to over 80% for longer SSRs. A corresponding increment was also observed for tri- and for tetra-meric SSRs.

In order to measure the informativeness of EST-derived barley SSR-markers, the polymorphism information content (PIC value) was calculated for each of the 76 mapped markers based on the six mapping parents used in this study. PIC values ranged from 0.28 to 0.78 with a mean PIC value of 0.45 ± 0.03. GBM1015 revealed the highest PIC value (0.78), which corresponds to the highest number of alleles (5) determined for this marker (Fig. 4). Interestingly, microsatellites consisting of the CCG motif, which were mainly situated in 5′-ESTs (84%), could hardly be transferred into polymorphic SSR-markers showing the predicted size (3%).

Linkage mapping

Of the 76 polymorphic markers 61 were mapped in OWBrec × OWBdom, 11 in the Steptoe × Morex and 16 in the Igri × Franka population. To support the development of a consensus map of the mapping populations, 12

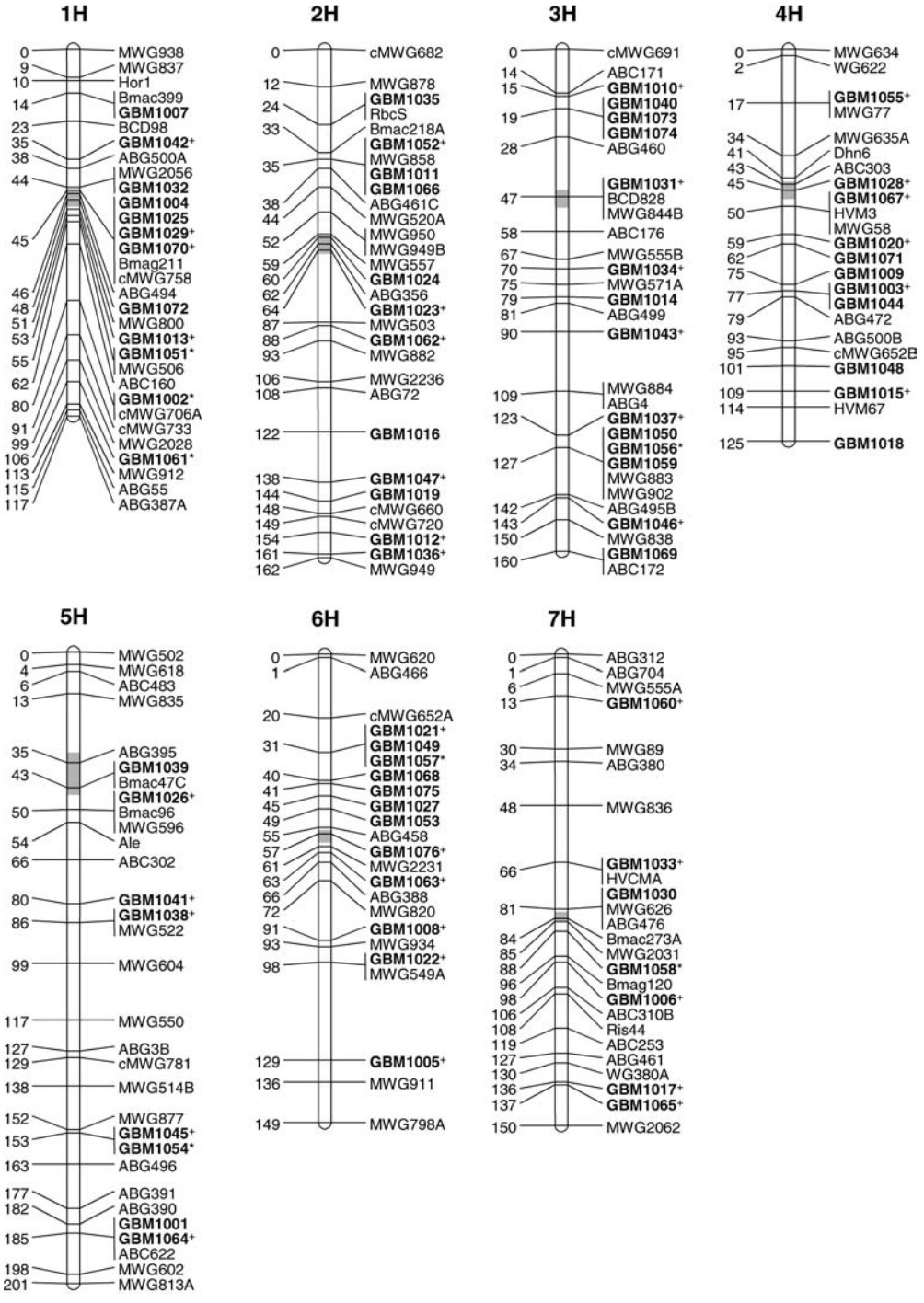

Fig. 5 A consensus genetic linkage map of barley. Skeleton markers used for map construction are shown in *black*. EST-derived SSR-marker loci are shown in *bold letters*. Map positions (in cM) are printed to the left. Centromeric regions are denoted by *grey boxes* (according to Künzel et al. 2000). * = marker that amplifies introns during PCR with genomic DNA; + = marker that was used for diversity analysis

SSRs were mapped in both I/F and OWB (there were enough markers in common between I/F and S/M). The SSR-markers were fairly even distributed with numbers ranging from 7 (7H) to 14 (3H) per chromosome. Clusters with three or more markers are present on chromosomes 6H, 3H, 2H and 1H, with the latter being located in the centromeric region of the chromosomes while the remaining clusters are located in more distal regions of the map (Fig. 5).

Co-segregating markers were re-checked regarding redundancy. As expected, GBM1011 and GBM1066 (2HS), derived from either the 5′-EST (HY05J06V) and the 3′-EST (HY05J06u) of the same cDNA clone, represent two different microsatellites that cosegregate. The co-segregating markers GBM1040 and GBM1073 (3HS), as well as GBM1050 and GBM1059 (3HL), appear to be derived from genes which are members of gene families, since the respective ESTs show homologies of 65% and 54%, respectively. Nevertheless, the sequences of the primer pairs are unique and should bind only to one target sequence under the PCR conditions applied in this study.

Table 4 Transferability of barley EST-derived SSR-markers to related species

Species/accession		Transferability[a]	
H. bulbosum	A17	36 (17)	77%
	2032	37 (16)	79%
	2920/4	37 (16)	79%
	2929/1	36 (15)	77%
	GRA951.80	41 (22)	87%
T. aestivum		29 (14)	62%
S. cereale	P87	29 (7)	62%
	P105	26 (5)	55%
O. sativa		19 (12)	40%

[a] Number of successfully amplified barley markers; relative frequencies in relation to the total number of markers investigated (47); figures in parenthesis stand for the number of markers for which PCR products were obtained being within the range of fragment sizes detected in 55 *H. vulgare* accessions

Interspecific transferability

A subset of 47 SSR-markers was used to assess the extent of transferability to related species. For this purpose five accessions of wild barley (*H. bulbosum*), two accessions of rye (*S. cereale*), and one accession each of wheat (*T. aestivum*) and rice (*O. sativa*) were analyzed. Only microsatellites were chosen which amplified a single product of the expected size in *H. vulgare*. In order to avoid unspecific amplification, the stringency of the PCR reaction was not reduced. In *H. bulbosum*, 1 to 4 amplicons were obtained during PCR. A relation between the ploidy level and the number of alleles was not observed (accessions A17 and GRA951.80 were tetraploid, the other three accessions were diploid). In the case of wheat, 1 to 3 amplicons were observed which is in good accordance with the existence of up to three orthologous loci within the A, B and D genome. In rye (diploid cross-pollinator) mostly one but sometimes also two amplicons were observed, while rice (diploid self-pollinator) consistently showed single PCR fragments. Approximately 80% of examined barley markers were successfully amplified in all the wild barley accessions, but accession GRA951.80 holds an exceptional position with an amplification rate of 87%. In case of this accession a high frequency of amplicons displaying the same fragment size than that of *H. vulgare* was noted. About 60% of the primers yielded amplification products in wheat and rye, whereas in rice the amplification rate decreased to 40% (Table 4). Of 25 markers that could be successfully transferred to rye, 11 (44%) were polymorphic between accessions 'P87' and 'P105', which are the parents of a mapping population (Korzun et al. 2001). Interestingly, sequencing of two SSR-markers amplified from rye DNA revealed a size polymorphism that is not caused by the SSR, but by an Insertion/Deletion of three nucleotides outside the microsatellite (GBM1047), or the absence of the microsatellite that was polymorphic in barley and several size polymorphisms in very short additional microsatellites, i.e. $(ATG)_{3/4}$ or $(CGG)_{2/3}$ (GBM1055).

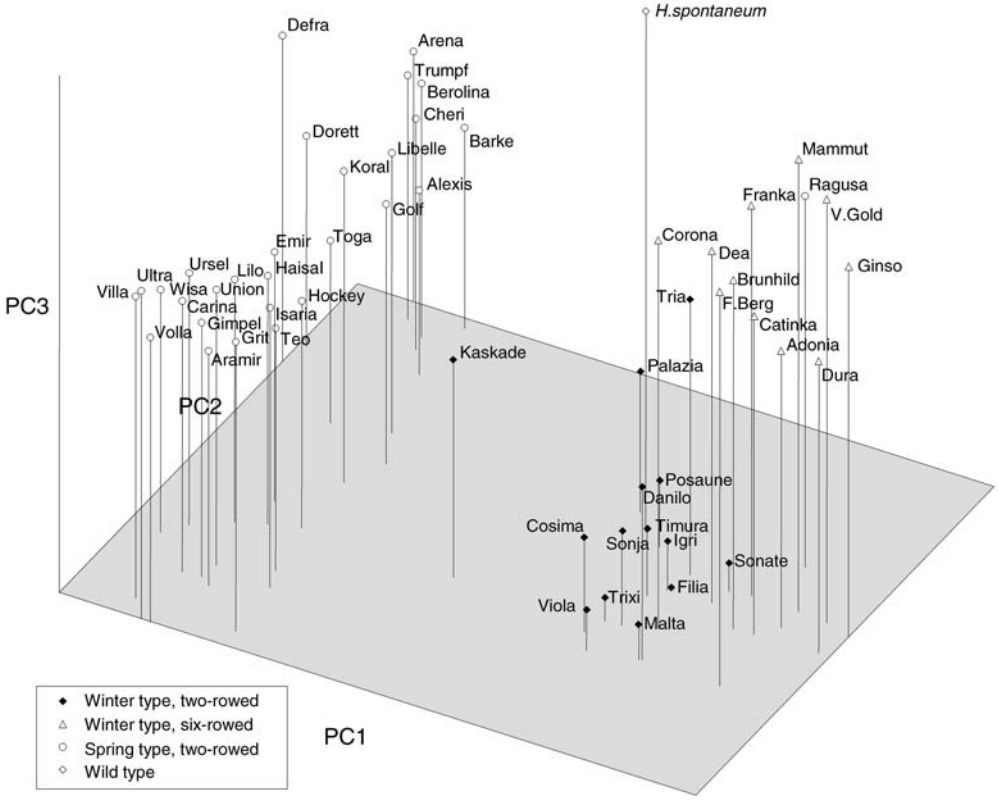

Fig. 6 Three-dimensional principle coordinate analysis (PCoA) plot of 54 barley cultivars, comprising 25 winter barley cultivars (11 two-rowed, 14 six-rowed) and 29 spring barley cultivars, and one *H. spontaneum* accession based on genetic similarity values deduced from 114 alleles derived from 38 EST-derived SSR-markers. Winter and spring barley as well as the wild form are depicted by different symbols. 'F.Berg' = 'Friedrichswerther Berg', 'V.Gold' = 'Vogelsanger Gold'; PC1, PC2, PC3 = first, second and third principal coordinates. PC1, PC2, and PC3 account for 25.4%, 10.0%, and 8.6% of the total variation

Diversity analysis

The usefulness of cDNA-derived SSR-markers for diversity analysis was investigated in a set of 54 barley cultivars and one additional *H. spontaneum* accession. A similarity matrix was calculated based on 114 alleles of 38 markers and the genetic relationships within this set of germplasm have been summarized by a PCoA presented in Fig. 6. The markers used in this study were evenly distributed throughout the genome, ranging from four markers on chromosome 1H, five markers on the chromosomes 5H and 7H and up to six markers on the chromosomes 2H, 3H, 4H and 6H (chromosomal positions of used markers are indicated in Fig. 5). The first three principal coordinates (PCs) accounted for 25.4%, 10.0% and 8.6% of the total variation in genetic similarity estimates, respectively.

PC1 separated the winter cultivars from the spring cultivars, while PC2 was mainly responsible for the wider spreading of the spring cultivars. PC3 divided the winter cultivars according to their spike type into six-rowed and two-rowed forms. PC3 was also responsible for the separation of the *H. spontaneum* accession from the remaining cultivars.

There were three bi-allelic markers showing largely disjoint patterns for the winter and spring cultivars, all of which mapped on chromosome 1H. GBM1042 was localized in the proximal regions of 1HS, and GBM1070 as well as GBM1029 mapped to the centromeric region. For the cultivars 'Ragusa' (spring type) and 'Kaskade' (winter type) the latter two markers, however, displayed the allele observed for the opposite barley type. In the case of GBM1042, the spring cultivars 'Barke', 'Ragusa' and 'Toga' revealed the allele which was observed for all winter cultivars.

Discussion

SSR frequency

To extend the available resource of barley SSR-markers a database comprising about 25,000 barley ESTs was systematically searched for SSR motifs. Electronic mining of the IPK EST database (B-EST, http://pgrc.ipk-gatersleben.de/) using a software tool (*MISA*) specifically developed for this purpose yielded a large set of candidate SSR-markers. A non-redundant set of SSR primers could be designed for 3.1% of the total number of ESTs analyzed. This figure is similar to those found in other species such as grape or sugarcane, where the frequency of non-redundant SSRs with dinucleotides of seven or more repeats and trinucleotides of five or more repeats was 2.5% and 2.88% of the total population of cDNA clones in the library, respectively (Scott et al. 2000; Cordeiro et al. 2001). Thus the frequency of unique microsatellite loci, where flanking primers could be successfully designed for, is well below the 30% that were observed in enriched genomic libraries (Ramsay et al. 2000). On the other hand, EST-derived SSR-markers are an almost free by-product of ongoing EST sequencing projects, entailing minimal costs for their identification.

In a comprehensive computational study of Cardle et al. (2000), average distances (in kb) between SSRs in sets of redundant ESTs were estimated as follows: rice (3.4), maize (8.1), soybean (7.4), tomato (11.1), poplar (14.0), *Arabidopsis* (13.8) and cotton (20.0). By applying the same criteria (NN repeats >14 bp, NNN repeats >15 bp, tetra- and penta-meric SSRs >4 repeats) SSRs occur with a frequency of 1/6.3 kb in barley ESTs. Hence, it seems that the frequency of cDNA-SSRs in the expressed portion of the barley genome is high in comparison to other plant species.

Among the dimeric repeats the motifs AG (52%) and AC (33%) were most common in our dataset, whereas AT microsatellites were detected only rarely (10%). The deficiency of AT SSRs in EST sequences is in accordance with reports from rice (Temnykh et al. 2000), *Arabidopsis* (Cardle et al. 2000) and maize (Chin 1996).

The dominance of trimeric SSRs in the present study can be explained by the suppression of non-trimeric SSRs in coding regions due to the risk of frameshift mutations that may occur when those microsatellites alternate in size of one unit (Metzgar et al. 2000). Among trimeric motifs CCG, AGG and AGC were the most apparent (65%). Similarly in rice, 60% of EST-derived microsatellite sequences were represented by the motifs CCG, ACG, AGG and ACC (Temnykh et al. 2000), while in maize CCG/GGC and AGG/CCT were most abundant (Chin 1996). CCG was also the most common motif in sugarcane (Cordeiro et al. 2001). Interestingly, CCG microsatellites appeared most frequently in 5′-ESTs (84%) in barley, whereas the remaining trimeric microsatellites are more or less equally distributed (56% in 5′-ESTs). In contrast, the motifs AAG and ATC represent 60% of all microsatellite motifs of the dicot *Arabidopsis* (Cardle et al. 2000). In barley, rice, maize, and sugarcane, as well as in *Arabidopsis*, the motif AAT appeared least often, probably because TAA-based variants code for stop codons that have a direct effect on protein synthesis (Chin 1996).

Marker development

Only 64% of the designed primer pairs proved to be functional. The remaining 36% primer pairs either failed or resulted in weak amplification. Similar findings were made for sugarcane (Cordeiro et al. 2001), where 40% of all primer pairs failed to amplify products. Possible explanations for this could be that primers extend across a splice site, the presence of large introns in the genomic sequence, the usage of questionable sequence information for primer development, or primers that were derived from chimeric cDNA clones. Certainly such failures can be minimized by rigorous quality checks of EST sequences and by careful cluster analysis, although

there is evidence that the quality of the sequence data used in this study is good (Michalek et al. 2002). To minimize primer failure due to low quality sequence reads, sequence information exceeding 700 bp was automatically rejected by the software used. Moreover, to provide a platform for high-throughput analyses, we deliberately chose only one single amplification profile and did not perform any further optimization step in order to increase the portion of useful primers for the sake of the applicability of the markers.

With regard to developing microsatellite markers, 3′-sequences yielded more polymorphic markers (64%) than 5′-ESTs did. This result is not unexpected. Due to the process of cDNA generation (polyT priming) there is a preferential selection of untranslated regions (UTR) within 3′-ESTs, therefore being more variable than 5′-ESTs. The expansion of non-trimeric microsatellites is suppressed in translated regions due to the risk of a frameshift mutation. This is supported by the fact that polymorphic microsatellites consisting of dimeric or tetrameric units could be found predominantly in 3′-ESTs (77% and 90%), but rarely in 5′-ESTs, whereas polymorphic trimeric microsatellites were found in nearly equal frequencies in 3′-ESTs (42%) compared to 5′-ESTs (58%).

Compared to genomic microsatellites, EST-derived microsatellites are generally shorter (22.7 ± 1.7 compared to 7.3 ± 1.1 repeats, Ramsay et al. 2000). Nevertheless, some exceptionally long microsatellites were identified with the largest SSR-marker (GBM1015) showing 19 repeats, $(ACAT)_{19}$, in cultivar 'Igri'. In the literature there is also an indication of long and extraordinarily polymorphic microsatellites in genes of barley, occurring in 28 and 37 alleles, respectively (Saghai Maroof et al. 1994). The mean PIC-value of genomic SSRs with 0.58 ± 0.03, estimated for eight spring, six winter cultivars and two wild barley accessions (Ramsay et al. 2000), is higher compared to the estimated mean PIC-value for cDNA-SSRs with 0.45 ± 0.03, based on the six mapping parents used in this study. These genotypes seem to satisfactorily represent the genetic variation present in the gene pool of cultivated barley, since the inclusion of a *H. spontaneum* line did not lead to a significant increase in the PIC values of a set of 75 EST-derived SSR markers (Kota et al., unpublished).

Given the growing number of barley ESTs available, the stringency of SSR selection can be increased, thus maximizing the number of highly informative markers. As can be seen from Fig. 3, a rise in the lower threshold of the number of repeat units to be considered for primer development can substantially increase the rate of polymorphic markers. Additionally, the presence of multiple SSRs raises the chance of obtaining polymorphic markers. The software developed allows the adjustment of thresholds for any type of repeat as well, as the choice of a specific parameter which facilitates the design of primer pairs flanking multiple SSRs.

Genetic mapping of 76 mapped EST-derived microsatellite loci resulted in some minor clusters which, however, are not concentrated around the centromeric region as was observed for a set of genomic microsatellites by Ramsay et al. (2000). This difference in distribution might be attributable to the sampling of different genomic regions by both types of markers. There is no evidence from the sequences available, that the cDNA-SSRs are associated with retroelements, which in turn might account for the formation of centromeric clusters of genomic SSRs. We rather expect that the distribution of the SSRs in the present map mirrors the distribution of genes along the genetic map of barley. Added value is being obtained from information on the putative function of the mapped SSRs. In this regard, 15 loci out of the 76 loci have already been annotated to putative gene functions via BlastX search. An additional seven loci showed significant homologies to proteins of unknown function from *Arabidopsis*.

Interspecific transferability

By virtue of the sequence conservation of transcribed regions of the genome a significant portion of the markers can be transferred to related *Triticeae* species. Transferability of EST-derived markers at different taxonomic levels has been demonstrated earlier (e.g. Scott et al. 2000; Cordeiro et al. 2001). Approximately 80% of the markers in the present study were transferred to *H. bulbosum* followed by rye, wheat (both 60%) and rice (40%). BlastN searches of 76 SSR loci against publicly available wheat and rice ESTs support the experimentally observed results, since 28 wheat ESTs, but less rice ESTs (11), showed significant homology to barley SSR loci (p-value < E-20). As expected, the proportion of transferable markers negatively correlates with the phylogenetic distance to barley: wild barley and cultivated barley belong to the same genus (*Hordeum*). Rye, wheat and barley are included in the same tribe *Triticeae* within the subfamily of the *Pooideae*. Rice belongs to the *Oryzeae* within the subfamily of the *Bambusoideae*. Also, it can be assumed that with increasing genetic distance, less-likely orthologous sequences will be detected. This will be of particular relevance for the comparisons between barley and rice. To get an idea of what portion of the sequences may detect orthologous loci in rice, the sequences of mapped barley SSRs were compared to the mapped rice ESTs. In this way six barley markers were identified that showed a significant homology (p-value < E-20). Of these three (50%) are located in syntenous positions in the two species (GBM1042, 1HS; GBM1051, 1HL; GBM1054, 5HL). On the other hand, 100% of the rice cDNAs-probes that detected single-copy fragments in barley mapped to a syntenous position on chromosome 3H (Smilde et al. 2001). Assuming that a successful cross species-hybridization requires sequence similiarity exceeding 70%, it can be assumed that, applying this (more stringent) criterion to cDNA-SSR-markers, the portion of orthologous loci can be further increased.

However, even if orthologous sequences are amplified, it must be kept in mind that even within the *Triticeae* the occurrence of variant alleles is not necessarily due to variability within the SSR. In this context, the rye amplicon of GBM1055 is characterized by a series of deletion and inversion events which had occurred outside the SSR, altogether yielding a net difference in fragment size between the two rye accessions of 2 bp, which is not due to a variation in the SSR sequence. This points to the limits of the interspecific transfer of SSR-markers for diversity studies, since (1) markers that show identical fragment sizes are not necessarily identical by descent (homoplasy), and (2) two alleles of different size might be the result of rather complex mutational events that mimic variation of the SSR. Therefore it remains uncertain if preservation of an individual SSR locus is more likely within the portion of markers which amplify fragments that are within the size range of fragments detected for *H. vulgare* accessions (see also Table 4).

Comparison of RFLPs and EST-derived SSRs for diversity studies

The application of EST-derived SSRs to analyze the genetic structure of 54 cultivars representative of the European barley genepool resulted in the differentiation of spring and winter types and, within the latter group, of two- and six- rowed forms. This result was not unexpected, since in a series of marker studies a similar differentiation has been observed (for a review see Graner et al. 2002). The genetic similarity matrix obtained from the SSR-markers was significantly correlated (Pearson coefficient r = 0.7) to the matrix obtained for the same set of cultivars using a set of RFLP markers, thus underscoring the usefulness of cDNA-SSRs for pedigree analysis (Melchinger et al. 1994). Given the fact that a large portion of RFLP markers were derived from transcribed regions of the barley genome, the over-all equivalence of both marker types is not surprising (Michalek et al. 1999).

Interestingly, in contrast to the previous RFLP data (Melchinger et al. 1994), the present SSR-marker data confirm the intermediate position of the cultivars ‘Ragusa’ and ‘Kaskade’ which reflect their position in various pedigrees. The two-rowed winter cultivar ‘Kaskade’ holds an exceptional position, since ‘Kaskade’ stands isolated between the spring and winter cultivars. This observation corresponds to the fact, that this cultivar has both spring (‘Union’, ‘Emir’) and winter cultivars (‘Malta’, ‘Carsten’) in its pedigree. Similarly, cultivar ‘Ragusa’, although being a spring type, groups among the winter cultivars. This can be explained by the fact that ‘Ragusa’ belongs to the principal progenitors of six-rowed winter barley cultivars, where it was systematically introgressed because of its mildew resistance (Fischbeck 1992).

Regarding the differentiation of spring and winter barley, three markers located on chromosome 1H showed a disjoint pattern. This observation is in accordance with a previous study based on the same data set, in which an RFLP marker that differentiated both germplasm groups was also assigned to chromosome 1H (Graner et al. 1990). Spring and winter barley differ in their requirement for vernalization, for which three loci are known. The spring growth habit loci *Sgh1* and *Sgh2* map to chromosomes 4HL and 5HL, while *Sgh3* has been assigned to the proximal region of barley chromosome 1H, which corresponds to the location of the markers (Takahashi and Yasuda 1971). The analysis of a larger set of European spring and winter barley germplasm will be required to confirm the observed association.

In conclusion, EST databases provide a valuable resource for the development of SSR-markers, which are associated with transcribed genes. Development of SSR-markers from EST-databases saves both cost and time, once a sufficient amount of EST sequences are available. The low level of polymorphism of EST-derived SSRs in comparison to genomic SSRs may be compensated for by their potential of interspecific transferability. In addition to mapping ESTs via microsatellite loci for locating the genes of putative functions, the EST-derived SSRs developed in this study are a useful tool to study the genetic diversity within cultivated barley, although the potential limitations imposed by homoplasy and allele mimicry require further attention.

Acknowledgements We are grateful to Patrick Hayes (Oregon State University) for providing the DH lines of barley mapping populations ‘Steptoe’ × ‘Morex’ and ‘OWBdom’ × ‘OWBrec’, to Richard A. Pickering (Crop and Food Research, Christchurch, New Zealand) for providing the four *H. bulbosum* accessions (‘A17’, ‘HB2032’, ‘Cb2920/4’ and ‘Cb2929/1’) and Andreas Börner for contributing the DNA of two rye inbred lines (‘P87’ and ‘P105’). We thank Winfriede Weschke and Elena Potokina for providing cDNA libraries, and Hangning Zhang for cleaning and editing the EST sequences. The present work was funded by grants from the State of Saxony-Anhalt (2686A/0087G), the Grains Research and Development Corporation (GRDC, UA 476), and the Federal Ministry of Education and Research (BMBF, 0312271C).

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. Genome 36:181–186

Baumer M, Cais R (2000) Abstammung der Gerstesorten. Bayerische Landesanstalt für Bodenkultur und Pflanzenbau, Freising, Germany

Becker J, Heun M (1995) Barley microsatellites: allele variation and mapping. Plant Mol Biol 27:835–845

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156:847–854

Chin ECL (1996) Maize simple repetitive DNA sequences: abundance and allele variation. Genome 39:866–873

Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. Plant Sci 160:1115–1123

Costa JM, Corey A, Hayes PM, Jobet C, Kleinhofs A, Kopisch-Obusch A, Kramer SF, Kudrna D, Li M, Riera-Lizarazu O, Sato K, Szucs P, Toojinda T, Vales MI, Wolfe RI (2001) Molecular mapping of the Oregon Wolfe barleys: a phenotypically polymorphic doubled-haploid population. Theor Appl Genet 103:415–424

Eujayl I, Sorrells ME, Baum M, Wolters P, Powell W (2002) Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor Appl Genet 104:399–407

Fischbeck G (1992) Barley cultivar development in Europe – success in the past and possible changes in the future. In: Munk L. Barley genetics VI. (ed) Mungaard International, Copenhagen, pp 885–901

Graner A, Siedler H, Jahoor A, Herrmann RG, Wenzel G (1990) Assessment of the degree and type of restriction fragment length polymorphism in barley (*Hordeum vulgare*). Theor Appl Genet 80:826–832

Graner A, Jahoor A, Schondelmaier J, Siedler H, Pillen K, Fischbeck G, Wenzel G, Herrmann RG (1991) Construction of an RFLP map of barley. Theor Appl Genet 83:250–256

Graner A, Bjørnstad Å, Konishi T, Ordon F (2002) Molecular diversity of the barley genome. In: Bothmer Rv, Hintum Tv, Knüpffer H, Sato K (eds) Diversity in barley (*Hordeum vulgare*). Elsevier, Amsterdam, The Netherlands (in press)

Hackauf B, Wehling P (2002) Identification of microsatellite polymorphisms in an expressed portion of the rye genome. Plant Breed 121:17–25

Kleinhofs A, Kilian A, Saghai Maroof MA, Biyashev RM, Hayes PM, Chen FQ, Lapitan N, Fenwick A, Blake T, Kanazin V, Ananiev E, Dahleen LS, Kudrna D, Bollinger J, Knapp SJ, Liu B, Sorrells MMH, Franckowiak J, Hoffmann DL, Skadsen R, Steffenson BJ (1993) A molecular, isozyme and morphological map of the barley (*Hordeum vulgare*) genome. Theor Appl Genet 86:705–712

Korzun V, Malyshev S, Voylokov AV, Börner A (2001) A genetic map of rye (*Secale cereale* L.) combining RFLP, isozyme, microsatellite and gene loci. Theor Appl Genet 102:709–717

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugen 12:172–175

Künzel G, Korzu L, Meister A (2000) Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints. Genetics 154:397–412

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Liu ZW, Biyashev RM, Saghai Maroof MA (1996) Development of simple sequence repeat DNA markers and their integration into a barley linkage map. Theor Appl Genet 93:869–876

Melchinger AE, Graner A, Singh M, Messmer MM (1994) Relationships among European barley germplasm. I. Genetic diversity among winter and spring cultivars revealed by RFLPs. Crop Sci 34:1191–1199

Metzgar D, Bytof J, Wills C (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA. Genome Res 10:72–80

Michalek W, Künzel G, Graner A (1999) Sequence analysis and gene identification in a set of mapped RFLP markers in barley (*Hordeum vulgare*). Genome 42:849–853

Michalek W, Weschke W, Pleissner KP, Graner A (2002) EST analysis in barley defines a unigene set comprising 4,000 genes. Theor Appl Genet 104:97–103

Miller RT, Christoffels AG, Gopalakrishnan C, Burke J, Ptitsyn AA, Broveak TR, Hide WA (1999) A comprehensive approach to clustering of the expressed human gene sequence: the sequence tag alignment and consensus knowledge base. Genome Res 9:1143–1155

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269–5273

Pillen K, Binder A, Kreuzkam B, Ramsay L, Waugh R, Förster J, Leon J (2000) Mapping new EMBL-derived barley microsatellites and their use in differentiating German barley cultivars. Theor Appl Genet 101:652–660

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1:215–222

Qi X, Stam P, Lindhout P (1996) Comparison and integration of four barley genetic maps. Genome 39:379–394

Ramsay L, Macaulay M, Ivanissevich dS, MacLean K, Cardle L, Fuller J, Edwards KJ, Tuvesson S, Morgante M, Massari A, Maestri E, Marmiroli N, Sjakste T, Ganal M, Powell W, Waugh R (2000) A simple sequence repeat-based linkage map of barley. Genetics 156:1997–2005

Rohlf FJ (1997) NTSYSpc, version 2.02. Exeter Software, Applied Biostatistics Incorporated

Rozen S, Skaletsky HJ (1998) Primer3. Code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html.

Saghai Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81:8014–8018

Saghai Maroof MA, Biyashev RM, Yang GP, Zhang Q, Allard RW (1994) Extraordinarily polymorphic microsatellite DNA in barley: species diversity, chromosomal locations, and population dynamics. Proc Natl Acad Sci USA 91:5466–5470

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

Smilde D, Haluskova J, Sasaki T, Graner A (2001) New evidence for synteny of rice chromosome 1 and barley chromosome 3H from rice expressed sequence tags. Genome 44:361–367

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant J 3:739–744

Struss P, Plieske J (1998) The use of microsatellite markers for detection of genetic diversity in barley populations. Theor Appl Genet 97:308–315

Takahashi R, Yasuda S (1971) Genetics of earliness and growth habit in barley. In: Nilan RA (ed) Barley genetics II. Washington State University Press, Pullman, USA, pp 388–408

Temnykh S, Park WD, Ayres N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T, McCouch SR (2000) Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:697–712

Weber JL (1990) Informativeness of human (dC–dA)*n*·(dG–dT)*n* polymorphisms. Genomics 7:524–530