Theor Appl Genet (2007) 114:1105–1116
DOI 10.1007/s00122-007-0504-6

ORIGINAL PAPER

Single nucleotide polymorphisms in rye (*Secale cereale* L.): discovery, frequency, and applications for genome mapping and diversity studies

R. K. Varshney · U. Beier · E. K. Khlestkina · R. Kota · V. Korzun · A. Graner · A. Börner

Received: 21 July 2006 / Accepted: 7 January 2007 / Published online: 8 March 2007

Abstract To elucidate the potential of single nucleotide polymorphism (SNP) markers in rye, a set of 48 barley EST (expressed sequence tag) primer pairs was employed to amplify from DNA prepared from five rye inbred lines. A total of 96 SNPs and 26 indels (insertion–deletions) were defined from the sequences of 14 of the resulting amplicons, giving an estimated frequency of 1 SNP per 58 bp and 1 indel per 214 bp in the rye transcriptome. A mean of 3.4 haplotypes per marker with a mean expected heterozygosity of 0.66 were observed. The nucleotide diversity index (π) was estimated to be in the range 0.0059–0.0530. To improve assay cost-effectiveness, 12 of the 14 SNPs were converted to a cleaved amplified polymorphic sequence (CAPS) format. The resulting 12 SNP loci mapped to chromosomes 1R, 3R, 4R, 5R, 6R, and 7R, at locations consistent with their known map positions in barley. SNP genotypic data were compared with genomic simple sequence repeat (SSR) and EST-derived SSR genotypic data collected from the same templates. This showed a broad equivalence with respect to genetic diversity between these different data types.

Communicated by A. Schulman.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-007-0504-6) contains supplementary material, which is available to authorized users.

R. K. Varshney · U. Beier · R. Kota · A. Graner · A. Börner
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstraße 3, 06466 Gatersleben, Germany

E. K. Khlestkina
Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Lavrentyeva Avenue 10, Novosibirsk 630090, Russia

V. Korzun
Lochow-Petkus GmbH, Grimsehlstraße 31, 37574 Einbeck, Germany

R. K. Varshney (✉)
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502 324 Andhra Pradesh, India
e-mail: r.k.varshney@cgiar.org

Present Address:
R. Kota
Plant Disease Resistance Group, CSIRO, Plant Industry, PO Box 1600, Canberra ACT 2601, Australia

Introduction

Single nucleotide polymorphisms (SNPs) are the most basic unit of genetic variation and represent the commonest class of DNA-based markers (Cho et al. 1999; Rafalski 2002). As a result, they can, in principle, be used to construct genetic maps with an at least 100-fold higher marker density than is possible using microsatellites (or simple sequence repeats—SSRs). The higher genetic stability of SNP over SSR, which is currently the most widely used marker platform in crop systems, is a further incentive for their development (Cho et al. 1999). In addition to providing an enhanced mapping resource, other applications, such as the assessment of genetic diversity, marker-assisted breeding, and the detection of genome-wide linkage disequilibrium and genotype/phenotype associations would benefit from their wide-scale development. All

these activities would become more practicable than is possible with SSRs, thanks to the ease with which SNP genotyping can be automated (Rafalski 2002). A major route for SNP discovery in genic sequence starts with an in silico comparison of homologous sequences from two or more representatives of a given species. Variants identified in this way generally need to be validated in vitro by resequencing, before specific SNP assays can be designed and tested. This broad approach has been employed for SNP discovery in rice (*Oryza sativa*; Nasu et al. 2002; Feltus et al. 2004), maize (*Zea mays*; Tenaillon et al. 2001; Ching et al. 2002), wheat (*Triticum aestivum*; Somers et al. 2003; http://www.wheat.pw.usda.gov/SNP/), barley (*Hordeum vulgare*; Kota et al. 2001; Russell et al. 2004; Rostoks et al. 2005), soybean (*Glycine max*; Zhu et al. 2003; Van et al. 2004), and sugarbeet (*Beta vulgaris*; Möhring et al. 2004).

Rye (*Secale cereale* L.) is a significant crop in Northern and Eastern Europe. Its comparative advantage over the other temperate cereals lies in its excellent tolerance to low temperature and high levels of soil aluminium, and its ability to realise acceptable grain-yield where other crops cannot (Madej 1996). Rye is also an important reservoir of genes for wheat improvement, and is a parent of triticale, the synthetic wheat–rye hybrid, which occupies a significant niche in the cropping system. As a result, the identification and genetic mapping of genes responsible for enhanced agronomic traits and abiotic stress tolerances is useful beyond their immediate benefit to rye genetics and breeding. To achieve these goals, well-saturated molecular genetic maps of rye are required. Several genetic linkage maps, constructed from a variety of populations and using various marker platforms have been developed for rye (Devos et al. 1993; Philipp et al. 1994; Senft and Wricke 1996; Korzun et al. 2001; Bednarek et al. 2003, summarized in Varshney et al. 2004 and Chikmawati et al. 2006). In fact, the exploitation of the heterologous RFLP (restriction fragment length polymorphism) markers, developed for wheat and barley, for constructing the genetic maps of rye have established good relationships among different linkage groups and genomes of three *Triticeae* species, i.e. wheat, barley, and rye (Devos et al. 1992, 1993; Devos and Gale 1993). However, in recent years, as with other crops, SSRs have come to represent the markers of choice for breeding applications (Gupta and Varshney 2000). In an attempt to increase the limited number of functional SSR assays in rye, Khlestkina et al. (2004, 2005) developed SSR markers from rye ESTs (expressed sequence tags) and also transferred the wheat-originating SSRs (Röder et al. 1998, 2004) to the rye genetic maps. They were able to map a number of both EST-derived (eSSR) and genomic (gSSRs) loci in four rye mapping populations. A particular problem in using such heterologous SSR assays is that only a small proportion of gSSR markers is transferable, and that eSSR markers, while readily transferable, tend to be relatively non-polymorphic (Varshney et al. 2005). Presently, a large number of ESTs are available for wheat and barley as compared to rye and a comprehensive resource of EST-based markers including SNPs have been developed for barley at Gatersleben. Therefore it is anticipated that the existing resource of barley EST-based markers could be used for developing the genic markers in rye for enhancing the density of the rye genetic maps as well as providing additional anchoring points between rye and barley genetic maps.

The present study was undertaken with the following objectives: (a) to assess the possibility of using barley genomic resources for undertaking SNP discovery in rye, (b) to assess SNP frequency and nucleotide diversity in the rye genome, (c) to develop affordable assays for SNP genotyping, (d) to integrate SNP markers into the rye genetic maps, and (e) to make a comparison of SNP with gSSR and eSSR markers in the context of diversity analysis.

Materials and methods

Plant materials and PCR

DNA was prepared, as per Khlestkina et al. (2004), from five inbred lines (N2, N6, N7, P87, and P105) used as parents of the four F_2 mapping populations described by Korzun et al. (2001) and Malyshev et al. (2003), and from 74 individuals selected from each of the four mapping populations. A set of 48 primer pairs was sampled from a collection of >200 EST-based barley SNP assays. The 48 loci mapped to all seven barley linkage groups, and were associated with a high nucleotide diversity index (π) (unpublished data). The barley amplicons were of mean size 300–450 bp. All PCR procedures followed those described elsewhere (Kota et al. 2001, 2003).

SNP discovery

Rye amplicons were sequenced in both forward and reverse orientation using big dye-terminator chemistry (Applied Biosystems, Foster City, CA, USA). Base calling was carried out using Phred (Ewing et al. 1998).

Raw sequence data were trimmed using a sliding window of 50 bp with a minimum average Phred score of 20, and filtered for a minimum length of 100 bp. ''Sequencher'' (Gene Codes Corporation, Ann Arbor, MI, USA) software was then used to generate contigs from the forward and the reverse sequences of each genotype with the parameters: (a) minimum match 85%, (b) minimum overlap 20 nt, and (c) assembly algorithm ''dirty data.'' The sequences were validated by a manual inspection of the trace files and edited where appropriate. Finally, the contigs were aligned with either GCG Pileup or ClustalW (Gribskov et al. 1984; Thompson et al. 1994), and putative SNPs differentiating the five rye-inbred lines were identified.

Polymorphism information content (PIC) and π

The polymorphism information content (PIC) was defined as per Nei (1987) as

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

where k is the total number of alleles and P is the frequency of the ith allele at a given locus.

The genetic variability was modelled by π, defined by the ratio K/L, where K is the average number of polymorphic nucleotide sites in a sequence of length L bp (Nei and Li 1979). The standard deviation of π was calculated according to Hartl and Clark (1997).

Conversion into cleaved amplified polymorphic sequences

Sequence alignments were loaded in FASTA format into the ''SNP2CAPS'' tool (http://pgrc.ipk-gatersleben.de/snp2caps/; Thiel et al. 2004), which employs the REBASE database (version 304, March 24, 2003), containing the recognition sequences of 235 non-isoschizomeric and commercially available restriction enzymes. Potentially informative restriction enzymes were validated in vitro, following methods detailed elsewhere (Thiel et al. 2004).

Linkage mapping

Single nucleotide polymorphisms were mapped in at least one of the four mapping populations. MAPMAKER 2.0 software (Lander et al. 1987) was used to assign SNPs into the framework map(s) at a LOD score of 3.0. cM distances were calculated by applying the Kosambi map-unit function (Kosambi 1944).

Functional annotation

Amplicon sequences containing a mappable SNP were tested against the non-redundant peptide (NR-PEP) database (Refseq-release 11, June 2005) using BLASTX2 (Altschul et al. 1990) with a threshold value 1E-10. These analyses were performed using Heidelberg Unix Sequence Analysis Resources at Deutsches Krebsforschungszentrum (DKFZ, German Cancer Research Centre), Heidelberg, Germany (http://genome.dkfz-heidelberg.de/).

Comparison of SNP and SSR markers

Single nucleotide polymorphism genotyping data were collected on the rye inbred lines in the present study. On the similar set of genotypes, we obtained additional genotyping data using a set of genomic SSRs (gSSRs) and EST-derived SSRs (eSSRs) (Khlestkina et al. 2004, 2005). The details on these SSR markers have been provided in Supplementary Table 1). For comparing the potential of EST-derived SNP (eSNP) markers of the present study with gSSRs and eSSRs, the genotyping data obtained for all three types of markers were coded into presence/absence matrices. Separate pairwise genetic similarity distance matrices among entries were assembled, based on eSNP, gSSR, and eSSR markers. Genetic similarity matrices obtained for each type of marker were compared using the Mantel (1967) test.

Results

SNP discovery and frequency in rye

All of the assays successfully amplified from a template of barley cv. Barke DNA, but only one third of the 48 primer pairs generated an adequate amplicon (a single, well-amplified product) in the expected size range from all the five rye templates. A further eight primer pairs were functional for at least three of the five templates. The rye and barley amplicons generated from these 24 primer pairs were sequenced. The remaining 24 primer pairs amplified a scorable 260–800 bp product from, at best, only one of the rye templates. Sequence data of sufficient quality for at least three of the ryes and the barley were obtained for 21 of the 24 amplicons. Sequence analysis for SNP discovery was performed only for the region where sequence data were obtained for at least three of the five inbred lines. In all, alignments over a mean of 397 bp per marker (range 220–762 bp) were examined, representing in all 5.55 kb of the rye

Table 1 Identification and analysis of SNPs

Marker	Primer sequence (5′–3′)	Sequence surveyed (bp)	Number of genotypes examined	Number of SNPs identified	Number of indels observed	Haplotypes based on SNPs	PIC (haplotypes)	Average pi (π)
GBS0131	F: AAGATACTCCACACCGACCG R: GGGTGGGGAACTTTGATCTC	220	4	6	8	4	0.750	0.0235
GBS0186	F: CAACTGCAGCTTATTCGGGAT R: ACCTTGGAGATTGGTCCCAC	485	5	6	–	5	0.800	0.0124
GBS0284	F: AAGATCGTGCATACGTCAACCA R: CATAAGTTATCGCCGTGGCAG	327	5	6	–	4	0.720	0.0159
GBS0360	F: CATGCCGAAGAACAAGGGTA R: GACTCCCTCGTTGAGGCG	265	5	3	–	2	0.480	0.0081
GBS0456	F: TCACTGCAATGCAGATCACG R: CGGGTACGAGGTGATCAAGAG	410	5	12	8	4	0.720	0.0300
GBS0461	F: CACCGTTGCTGACACTGGAT R: AATGCGGCTCTTTGTGGG	405	5	7	–	4	0.720	0.0151
GBS0524	F: TGCCAGTTTAGCATCAATTTGC R: TTTTACCCACGTGAGAAGCTTG	398	3	1	–	2	0.667	0.0084
GBS0526	F: AGACAGAATCCTCACAGGTGCC R: CCATGCCGAAGCAGATCC	674	3	5	5	3	0.667	0.0196
GBS0551	F: GTGCAGCCTTGCCTTCATAA R: CGTCGGATTCAACGTCTCCA	762	5	29	2	5	0.800	0.0530
GBS0554	F: ATGGAGCCCCTCCCAACTAC R: GTAGACGTCCAGCACCTCGAT	478	2	5	2	2	0.500	0.0377
GBS0577	F: GTGCTCAACAATGCCCCCTA R: CAGGTTCTTGGCTGCTTGTATC	254	4	3	–	2	0.375	0.0059
GBS0582	F: CTGGAGAAACCAGCCTATGGA R: CCAGGCAATGCTCATGAATG	344	5	7	1	5	0.800	0.0174
GBS0613	F: AGTGTTACATGCATCGCACCG R: GGCTTCATCGTCTACCCTTCG	242	4	5	–	3	0.625	0.0186
GBS0712	F: TACGAAACTCTTGCTCGGGC R: CGGGCATACTCAGGCAAAG	290	4	1	–	2	0.625	0.0190
Total		5,554		96	26			
Average		396.71	4.21			3.36	0.661	0.0203

F forward primer, *R* reverse primer

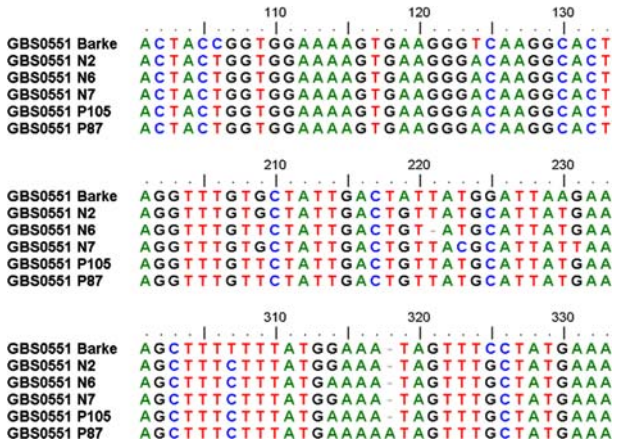

Fig. 1 Sequence alignment showing SNP discovery. Sequence alignment for GBS0551 across five rye inbred lines (N2, N6, N7, P87, and P105) and barley cv. Barke. A number of SNPs varying amongst the rye lines are identifiable

genome (Table 1). In summary, these results suggested the utility of existing resource of barley EST markers for undertaking SNP discovery in rye.

A total of 96 SNPs and 26 indels were observed in 14 of the 21 amplicons (Table 1; Fig. 1). The remaining seven amplicon-sequences were completely monomorphic across the five ryes and the barley. The highest number of SNPs per amplicon (29) was in GBS0551, followed by 12 in GBS0456, while indel frequency was highest (12) in GBS0131 and GBS0456. GBS0524 and GBS0712 yielded only one SNP in each, and eight of the amplicons lacked any indel. The overall frequencies of SNPs and indels were, respectively, 1 per 58 bp and 1 per 214 bp. π ranged from 0.0059 to 0.0530, with a mean of 0.0203.

Expected heterozygosity and haplotypes in rye

Single nucleotide polymorphism markers are usually biallelic and so, their PIC cannot exceed 0.50. However, by considering haplotypes (the combination of SNPs within an amplicon), higher PIC values can be generated. Between two and five (mean 3.36) haplotypes per amplicon were identified, resulting in haplotype-based PIC values lying between 0.37 and 0.80 (mean 0.66) (Table 1). Over 85% of the amplicons yielded a haplotype-PIC of >0.50. In contrast, the PIC values of individual SNPs fell in the range 0.22–0.50 (mean 0.32).

Comparison of sequence diversity between rye and barley

The set of markers analysed for sequence diversity is a subset of the barley EST markers that were analysed for sequence diversity in barley (R. Kota et al., unpublished). Screening of seven barley genotypes (Igri, Franka, Steptoe, Morex, OWBDom, OWBRec, and Barke), the parental genotypes of four mapping populations (i.e. Igri × Franka, Steptoe × Morex, OWBDom × OWBRec, and Barke × Morex), with these 14 markers yielded a total of 76 SNPs which is 1.26× lesser as compared to that was observed in rye (Table 2). Although these markers did not show any indel in barley genotypes, a total of 26 indels were observed in rye genotypes. Number of haplotypes observed in barley genotypes was a bit higher as compared to that of rye. The PIC value of the haplotypes and sequence diversity index in rye genotypes for the examined markers, however, was higher in rye as compared to that of barley (Table 2).

Conversion of SNPs into cleaved amplified polymorphic sequence markers for rye

Many SNP detection and genotyping platforms depend on expensive equipment and/or consumables, making SNP genotyping an expensive process. Of the 14 polymorphic sequence alignments, 12 contained at least one potential cleaved amplified polymorphic sequence (CAPS) candidate (Table 3). All these candidates were validated by the appropriate digestion reactions (Fig. 2).

Table 2 Comparative assessment of sequence diversity between rye and barley

Feature	Rye	Barley	Fold difference in rye as compared to barley
Genotypes analysed	5	7	
Number of SNPs identified	96 (6.85)	76 (5.42)	1.26× higher
Number of indels observed	26	0	New feature observed
Number of haplotypes per marker	2–5 (3.36)	2–6 (3.5)	0.96× less
PIC of haplotypes	0.38–0.80 (0.66)	0.22–0.81 (0.57)	1.16× higher
Average nucleotide diversity (π)	0.0059–0.0530 (0.0203)	0.0012–0.0150 (0.0052)	3.90× higher
Linkage groups represented	6 (not 2R)	6 (not 6H)	No change

Table 3 Mapped SNP loci in rye: location, CAPS conversion, and putative function

Rye linkage groups, marker loci mapped	Rye population used for mapping	Restriction enzyme for CAPS assay	Marker name and linkage groups in barley	Putative function (BLASTX description)	Protein ID[a]	E-value	E-score
1R							
Xgbs0131-1R	N6 × N2	*Mse*I	GBS0131 (1H)	MCB1 protein (*Hordeum vulgare*)	emb|CAC24844.1|	1E-109	397
Xgbs0554-1R	N7 × N6	*Hha*I	GBS0554 (1H)	Pathogenesis related protein (*H. vulgare*)	emb|CAA34641.1|	4E-48	191
3R							
Xgbs0186-3R	N7 × N2	*Hha*I	GBS0186 (3H)	Putative aspartate aminotransferase (*Oryza sativa*)	dbj|BAD87343.1|	7E-31	135
Xgbs0284-3R	N7 × N2	*Cac*8I	GBS0284 (3H)	Unknown protein (*O. sativa*)	dbj|BAD82692.1|	1E-16	88.2
4R							
Xgbs0456-4R	N7 × N2	*Rsa*I	GBS0456 (4H)	Glutamine synthetase isoform GSe1 (*Triticum aestivum*)	gb|AAR84349.1|	5E-54	211
Xgbs0551-4R	N7 × N2	*Apo*I	GBS0551 (4H)	RNA binding protein Rp120 (*O. sativa*)	gb|AAP85378.1|	5E-85	315
5R							
Xgbs0577-5R	N7 × N2	*Eco*RV	GBS0577 (5H)	Hordoindoline-a (*H. vulgare*)	gb|AAV37976.1|	7E-68	258
Xgbs0613-5R	P105 × P87	*Nci*I	GBS0613 (5H)	Putative cysteine conjugate beta-lyase (*O. sativa*)	dbj|BAD38029.1|	4E-53	209
Xgbs0712-5R	P105 × P87	*Dde*I	GBS0712 (5H)	NPH3 family protein (*O. sativa*)	gb|AAT85278.1|	2E-86	320
6R							
Xgbs0526-6R	N7 × N2	*Taq*I	GBS0526 (3H)	Putative 60S ribosomal protein L38 (*O. sativa*)	gb|AAT07599.1|	7E-32	137
7R							
Xgbs0461-7R	N7 × N2	*Nsp*I	GBS0461 (4H)	Iron-deficiency induced gene (*H. vulgare*)	dbj|BAB61039.1|	2E-22	106
Xgbs0360-7R	N7 × N2	*Eco*RV	GBS0360 (7H)	Eukaryotic translation initiation factor 1A (*O. sativa*)	emb|CAD91551.1|	3E-54	213

emb European Molecular Biology Laboratories (EMBL), *gb* Genbank, *dbj* DNA Databank of Japan

[a] Protein databases

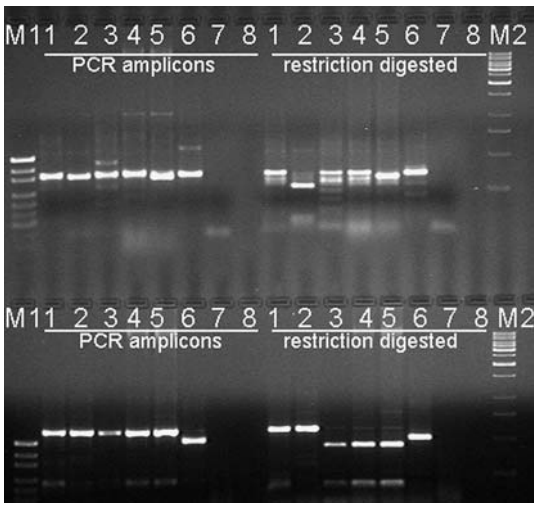

Fig. 2 Development of CAPS assays. *Mse*I and *Hha*I digests of two amplicons (*upper panel*: GBS0131, *lower panel*: GBS0186). The *left side* of each panel show undigested PCR products, and the *right side* digested PCR products. *M1* puC19 DNA/*Msp*I, size standard; *M2* 1 kb ladder, size standard; *1* N2; *2* N6; *3* N7; *4* P87; *5* P105; *6* Barke; *7, 8* water

Integration of SNP markers into rye genetic maps

The 12 CAPS assays were applied to individual progeny of the mapping populations, enabling the mapping of eight SNP loci in N7 × N2, two in P105 × P87, and one each in N6 × N2 and N7 × N6. All segregations were in accordance with the 1:2:1 ratio expected for an F_2 population. The loci were thus straightforwardly integrated into the pre-existing genetic maps of Khlestkina et al. (2004, 2005), thereby adding 12 loci, spread over all the rye chromosomes except for 2R (Table 3; Fig. 3). The majority of the loci mapped to the rye linkage group corresponding to the one in which the sequence was located in barley. The exceptions were GBS0526 (3H) and GBS0461 (4H), which were assigned to, respectively, chromosomes 6R and 7R (Table 3; Fig. 3).

Comparison of SNP with gSSR and eSSR markers in rye

The potential of SNPs to assess genetic diversity in rye was assessed by comparing the amplicon sequences obtained by 14 EST markers across the five inbred lines. This set of genotypic data (involving 96 SNPs) was used to describe the genetic relationships between the inbred lines (Fig. 4). The resulting dendrogram suggests that N2 is more diverse than the other lines, while P87 and P105 have a level of similarity of greater than 50%. A comparison was then made between the SNP genetic similarity matrix and those derived from 39 eSSR (114 alleles) and 60 gSSR (167 alleles) loci (Khlestkina et al. 2004, 2005; Supplementary Table 1). Analysis showed a broad equivalence with respect to genetic diversity between these different data types (Table 4; Fig. 4). A high-correlation coefficient was obtained between both the genetic diversity indices obtained from gSSR and eSSR data ($r = 0.992$), eSSR and SNP data ($r = 0.982$), and gSSR and SNP data ($r = 0.972$).

Discussion

Sequence diversity and SNPs in rye using barley markers

We have illustrated how eSNP markers in barley can be used to discover SNP variation in rye, and that these assays can be used to analyse sequence diversity in the rye genome. The possibility of transferring pre-existing barley markers to rye saves substantial time and effort, otherwise needed for the de novo design and synthesis of SNP primer pairs. Intergeneric transfer of DNA markers cannot always be relied upon, although the use of heterologous cDNA probes for RFLP analysis in the cereals has worked very well (Devos et al. 1992, 1993; Devos and Gale 1993; Nelson et al. 1995). Some success has also been achieved in transferring SSR markers from wheat and barley to rye (Khlestkina et al. 2004, 2005), especially to those based on genic, as opposed to anonymous genomic sequence (Gupta et al. 2003; Thiel et al. 2003; Varshney et al. 2005). This higher level of transferability no doubt reflects the better conservation over speciation of coding, as opposed to non-coding sequence (Varshney et al. 2005).

Polymorphism in allogamous species is generally higher than in autogamous ones (Rafalski 2002), and thus, as expected, SNP frequency in rye was higher than in barley (ranging between 1/78 and 1/189 bp; Kanazin et al. 2002; Bundock et al. 2003; Russell et al. 2004; R. Kota et al., unpublished), wheat (1/540 bp, Somers et al. 2003), sorghum (1/123 bp, Hamblin et al. 2004), sugarbeet (1/130 bp, Schneider et al. 2001), and soybean (1/278 bp, Van et al. 2004), but comparable to its frequency in maize (1/61 bp, Ching et al. 2002. Likewise, the mean π in rye is higher than that in barley (3.2×10^{-3}), wheat (6.9×10^{-3}, Somers et al. 2003), sorghum (2.25×10^{-3}, Hamblin et al. 2004), soybean (0.97×10^{-3}, Zhu et al. 2003; 0.70×10^{-3}, Van et al. 2004), and sugarbeet (7.6×10^{-3}, Schneider et al. 2001). Surprisingly, it was also two- to threefolds higher than that in maize (9.6×10^{-3}, Tenaillon et al. 2001; 6.3×10^{-3}, Ching et al. 2002). We believe,

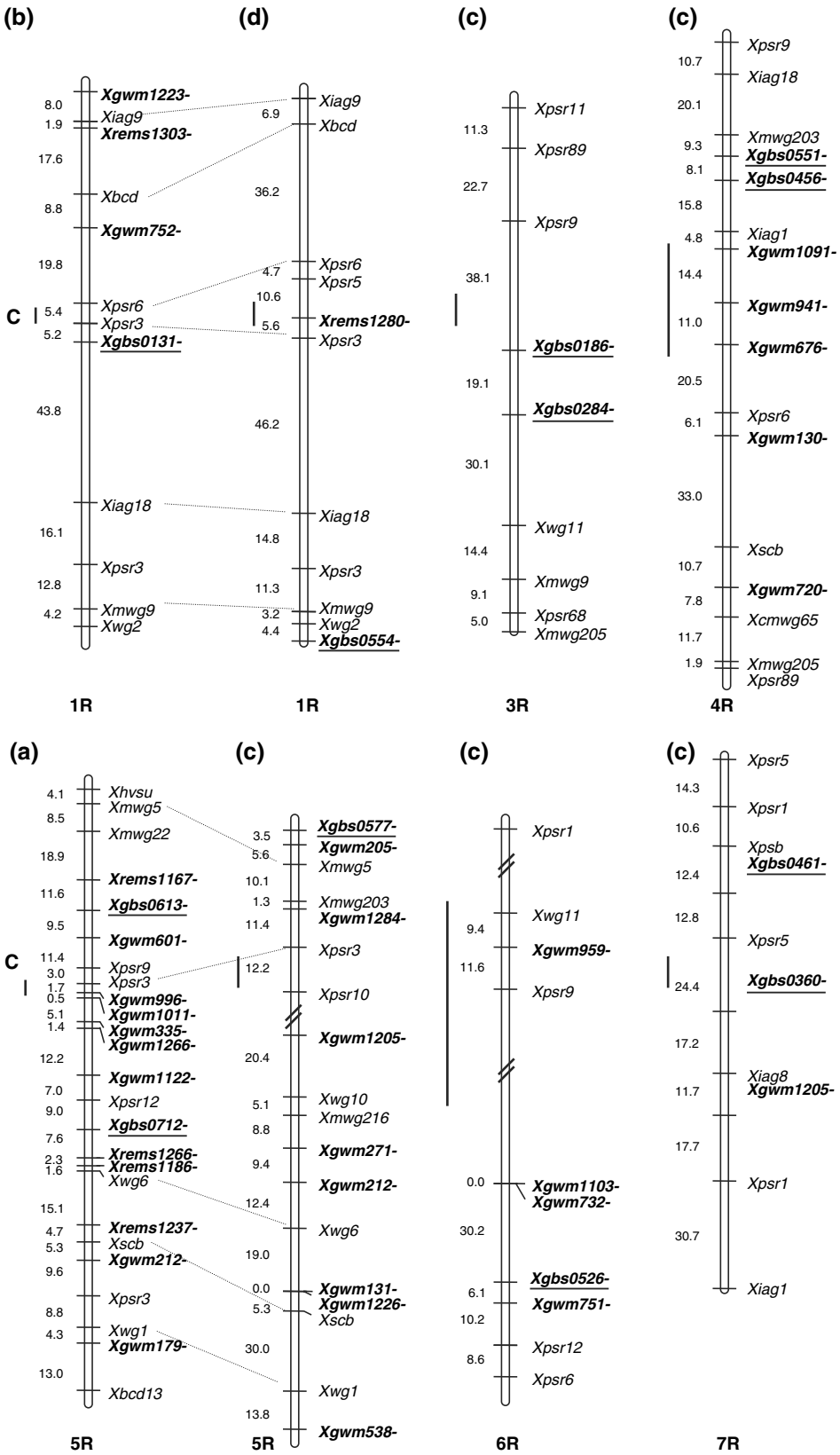

Fig. 3 Integration of SNP loci into the rye genetic map. SNP markers (*underlined*) integrated into rye microsatellite linkage map (Khlestkina et al. 2004) based on segregation data from progeny of crosses **a** P87 × P105, **b** N6 × N2, **c** N7 × N2, and **d** N7 × N6. The tentative position of centromeres is indicated based on data of Devos et al. (1993) and Korzun et al. (2001). Short arms of chromosomes are at the *top*, and the long arms at the *bottom*

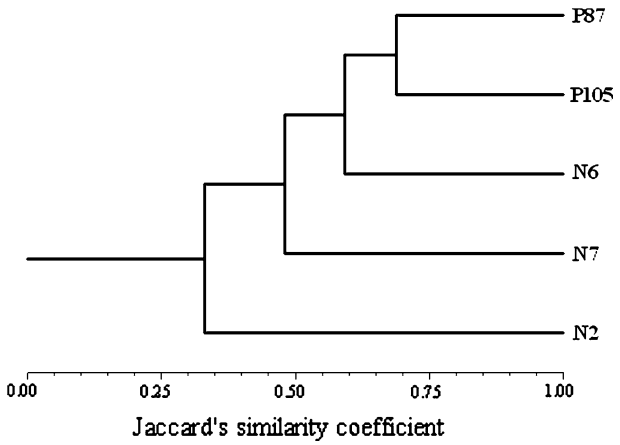

Fig. 4 Genetic relationships among five inbred lines defined by 96 SNPs obtained by using 14 EST markers

Table 4 Genetic diversity indices among rye genotypes based on SNP, eSSR and gSSR marker data

Marker type	P87	P105	N2	N6	N7	Genotype
SNP		0.688	0.385	0.643	0.503	
eSSR	1.000	0.565	0.608	0.549	0.480	P87
gSSR		0.465	0.520	0.560	0.512	
SNP			0.434	0.542	0.414	
eSSR		1.000	0.618	0.637	0.569	P105
gSSR			0.584	0.526	0.448	
SNP				0.313	0.192	
eSSR			1.000	0.556	0.539	N2
gSSR				0.512	0.486	
SNP					0.522	
eSSR				1.000	0.598	N6
gSSR					0.519	
SNP						
eSSR					1.000	N7
gSSR						

Allelic diversity data obtained for 96 SNPs with 14 SNP markers, 114 alleles with 39 eSSRs and 167 alleles with 60 gSSRs were used for comparison by using Jaccard's similarity coefficient

however, that this result reflects the non-random choice applied for the amplicons to screen, where a deliberate attempt was made to select markers with both high PIC and high π.

On the basis of allelic frequencies, the SNPs identified had a mean PIC of 0.32. In contrast, the gSSR and eSSR markers across the same germplasm sample had a mean PIC of, respectively, 0.55 and 0.52. However, when PIC was calculated on the basis of haplotype instead of on individual SNPs, its value was more than doubled. Several studies (e.g. Johnson et al. 2001) have suggested that haplotype analysis is superior to SNP analysis for trait association or diagnostic studies. Overall, the information content of SNP haplotypes is comparable or even higher than that of the SSR markers.

Higher sequence diversity in rye as compared to barley

As mentioned earlier, occurrence of indels, the higher number of SNPs (1.26×), higher sequence diversity (3.9×), and higher PIC value of haplotypes (1.16×) obtained in rye as compared to barley with the same set of the markers can be attributed to the allogamous nature of rye while the barley is an autogamous species (Rafalski 2002). A slightly higher number of haplotypes were observed in barley as compared to rye. This could be possible as a higher number of barley genotypes (seven) than rye (five) were analysed with these markers. In brief, these analyses clearly underline the value of existing resource of barley markers for undertaking SNP discovery and marker development studies in rye.

Development of functional SNP markers for rye

The present study represents the beginning of SNP discovery and development in rye. A major barrier for SNP discovery in rye to date has been the paucity of EST sequence and the poor level of genotype representation among these ESTs. This has militated against a database mining strategy for SNP discovery. We have shown that heterologous EST sequence can substitute for the lack of rye sequence. Following this discovery, the next challenge in SNP development is in the design of genotyping assays. Numerous competing platforms are being promoted for this purpose, but some are associated with high costs for specialized equipment and/or reagents. The conversion of SNPs to CAPS markers provides an opportunity for applications in laboratories equipped with only basic infrastructural facilities. Conversion was achieved here for 12 markers, allowing them to be deployed using simple PCR reactions, restriction digestion, and agarose gel electrophoresis in a very cost- and time-effective manner. Our aim was to demonstrate a route to the development of low cost SNP assays in rye, which would open the way to a substantial enrichment of the rye genetic map. This is particularly relevant for rye, for which, at present, the genetic maps in the public domain involve RFLP, AFLP, RAPD, and SSR loci (Devos et al. 1993; Philipp et al. 1994; Senft and Wricke 1996; Korzun et al. 2001; Bednarek et al. 2003; Khlestkina et al. 2004, 2005). None of these maps have a significant marker density (Varshney et al. 2004).

Of the 12 mapped SNP loci, 10 were located to their expected chromosome, based on their location in barley. However, the remaining two (GBS0526 and GBS0461, on barley 3H and 4H) mapped to apparently

non-homoeologous locations in rye (6R and 7R, respectively). These two rye chromosomes have evolved, via multiple translocations, to become substantially differentiated from other *Triticeae* species (Devos et al. 1993; Devos and Gale 1993), and these rearrangements are consistent with the apparently non-homoeologous locations of these two markers (Table 3).

As the source sequence for the EST-SNP markers is genic, putative function can often be assigned using standard homology searches within non-redundant protein databases. Except for one marker (*Xgbs0284-3R*), all sequences showed a level of homology with identified proteins (Table 3). A known function can give added value to a marker, since the possibility exists that SNPs can then be directly associated with variants for a specific function (Holton et al. 2002; Gao et al. 2004). Thus, GBS0554 (1R) and GBS0461 (7R) may be associated with quantitative variation for biotic and abiotic stress tolerance, as their sequences belong to a gene encoding a pathogenesis-related protein (GBS0554), or an iron-deficiency-induced protein (GBS0461). Were such an association to be confirmed through a genetic analysis, such markers would become useful both for marker-assisted selection and for allele mining in germplasm collections. Such a strategy requires a large repertoire of mapped genes (Andersen and Lübberstedt 2003).

Genetic diversity and comparison of SNP markers with gSSRs and eSSRs in rye

The relationships among the five inbred lines suggest that N2 is genetically distant from the other four inbred lines, while P87 and P105 are closely related to one another. This is consistent with known pedigree and provenance (Korzun et al. 2001; Malyshev et al. 2003; Tikhenko et al. 2005). P87 and P105 were developed together at the Institute of Genetics and Cytology, Minsk, Belarus, both having been selected from the progeny of a single cross. N6 and N7 originate from, respectively, Sweden and Russia, while N2 is thought to have been developed in Northern/Eastern Europe (N. D. Tikhenko, personal communication).

The SNP diversity is consistent with that derived from analysis of gSSRs and eSSRs. Importantly, even though only a small number of SNPs were analysed, highly significant correlations were obtained between the SNP marker data set and that of the gSSR ($r = 0.972$) and eSSR ($r = 0.982$) sets. The even higher correlation ($r = 0.992$) between the gSSR and eSSR data sets probably reflects the larger number of alleles samples. Since both eSSR and SNP markers are sourced from the transcriptome, it is not surprising that the correlation between them is so high.

In conclusion, we have shown that extant barley sequence can be used for SNP discovery in rye and that most of these SNPs can be converted into economical CAPS assays. We have used these new markers to add 12 functional loci to the rye map, and showed that these loci are syntenic between rye and barley. The development of CAPS assay should allow SNP markers to be deployed in situations where sophisticated infrastructure is lacking. The availability of a large number of ESTs in barley and wheat can facilitate the development of SNP markers, useful for both rye genetics and breeding.

Acknowledgment The authors thank Dr. Nils Stein, IPK, for his constructive suggestions during the course of study.

References

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

Andersen JR, Lübberstedt T (2003) Functional markers in plants. Trends Plant Sci 8:554–560

Bednarek PT, Masojc P, Lewandowska R, Myskow B (2003) Saturating rye genetic map with amplified fragment length polymorphism (AFLP) and random amplified polymorphic DNA (RAPD) markers. J Appl Genet 44:21–33

Bundock PC, Christopher JT, Eggler P, Ablett G, Henry RJ, Holton TA (2003) Single nucleotide polymorphisms in cytochrome P450 genes from barley. Theor Appl Genet 106:676–682

Chikmawati T, Ma XF, Ross K, Miftahudin, Gustafson JP (2006) Rye. In: Kole C (ed) The genomes: a series on genome mapping, molecular breeding in plants: cereals and millets, vol 1. Springer, Germany, pp 243–255

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 3:19

Cho RJ, Mindrinos M, Richards DR, Sapolsky RJ, Anderson M, et al (1999) Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. Nat Genet 23:203–207

Devos KM, Gale MD (1993) Extended genetic maps of the homoeologous group-3 chromosomes of wheat, rye and barley. Theor Appl Genet 85:649–652

Devos KM, Atkinson MD, Chinoy CN, Liu CJ, Gale MD (1992) RFLP-based genetic map of the homoeologous group 3 chromosomes of wheat and rye. Theor Appl Genet 83:931–939

Devos KM, Atkinson MD, Chinoy CN, Francis HA, Harcourt RL, Koebner RMD, Liu CJ, Masojc P, Xie DX, Gale MD (1993) Chromosomal rearrangements in the rye genome relative to that of wheat. Theor Appl Genet 85:673–680

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8:175–185

Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N, Paterson AH (2004) An SNP resource for rice genetics and breeding based on subspecies *indica* and *japonica* genome alignments. Genome Res 14:1812–1819

Gao LF, Jing RL, Hu NX, Li XP, Zhou RH, Chang XP, Tang JF, Ma Zy, Jia JZ (2004) One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat. Theor Appl Genet 108:1392–1400

Gribskov M, Devereux J, Burgess RR (1984) The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression. Nucleic Acids Res 12:539–549

Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Genet Genomics 270:315–323

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113:163–165

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

Hamblin MT, Mitchell SE, White GM, Gallego J, Kukatla R, Wing RA, Paterson AH, Kresovich S (2004) Comparative population genetics of the panicoid grasses: sequence polymorphism, linkage disequilibrium and selection in a diverse sample of *Sorghum bicolor*. Genetics 167:471–483

Hartl DL, Clark AG (1997) Principles of population genetics. Sinauer Associates, Sunderland

Holton TA, Christopher JT, McClure L, Harker N, Henry RJ (2002) Identification and mapping of polymorphic SSR markers from expressed gene sequences of barley and wheat. Mol Breed 9:63–71

Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA (2001) Haplotype tagging for the identification of common disease genes. Nat Genet 29:233–237

Kanazin V, Talbert H, See D, Decamp P, Nevo E, Blake T (2002) Discovery and assay of single nucleotide polymorphism in barley (*Hordeum vulgare*). Plant Mol Biol 48:529–537

Khlestkina EK, Than MH, Pestsova EG, Roder MS, Malyshev SV, Korzun V, Börner A (2004) Mapping of 99 new microsatellite-derived loci in rye (*Secale cereale* L.) including 39 expressed sequence tags. Theor Appl Genet 109:725–732

Khlestkina EK, Than MH, Pestsova EG, Roder MS, Malyshev SV, Korzun V, Börner A (2005) Mapping of 99 new microsatellite-derived loci in rye (*Secale cereale* L.) including 39 expressed sequence tags. Erratum. Theor Appl Genet 110:990–991

Korzun V, Malyshev S, Voylokov AV, Börner A (2001) A genetic map of rye (*Secale cereale* L.) combining RFLP, isozyme, protein, 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

Kota R, Varshney RK, Thiel T, Dehmer K-J, Graner A (2001) Generation and comparison of EST-derived SSR and SNP markers in barley (*Hordeum vulgare* L.). Hereditas 135:141–151

Kota R, Rudd S, Facius A, Kolesov G, Thiel T, Zhang H, Stein N, Mayer K, Graner A (2003) Snipping polymorphisms from large EST collections in barley (*Hordeum vulgare* L.). Mol Genet Genomics 270:24–33

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

Madej LJ (1996) Worldwide trends in rye growing and breeding. Vortr Pflanzenüecht 35:1–6

Malyshev SV, Kartel NA, Voylokov AV, Korzun V, Börner A (2003) Comparative analysis of QTLs affecting agronomical traits in rye and wheat. EWAC Newsl 12:120–122

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220

Möhring S, Salamini F, Schneider K (2004) Multiplexed, linkage group—specific marker sets for rapid genetic mapping and fingerprinting of sugar beet (*Beta vulgaris* L.). Mol Breed 14:475–488

Nasu S, Suzuki J, Ohta R, Hasegawa K, Yui R, Kitazawa N, Monna L, Minobe Y (2002) Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. DNA Res 9:163–171

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

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

Nelson JC, Sorrells ME, Van Deynze AE, Lu YH, Atkinson M, Bernard M, Leroy P, Faris JD, Anderson JA (1995) Molecular mapping of wheat: major genes and rearrangements in homoeologous groups 4, 5, and 7. Genetics 141:721–731

Philipp U, Wehling P, Wricke G (1994) A linkage map of rye. Theor Appl Genet 88:243–248

Polakova K, Laurie DA, Vaculova K, Ovesna J (2003) Characterization of beta-amylase alleles in 79 barley varieties with pyrosequencing. Plant Mol Biol Rep 21:439–447

Rafalski JA (2002) Application of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100

Röder MS, Huang X-Q, Ganal MW (2004) Wheat microsatellites: potential and implications. In: Lörz H, Wenzel G (eds) Biotechnology in agriculture and forestry, molecular marker systems, vol 55. Springer, Heidelberg, pp 255–266

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

Rostoks N, Mudie S, Cardle L, Russell J, Ramsay L, Booth A, Svensson JT, Wanamaker SI, Walia H, Rodriguez EM, Hedley PE, Liu H, Morris J, Close TJ, Marshall DF, Waugh R (2005) Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. Mol Genet Genomics 274:515–527

Russell J, Booth A, Fuller J, Harrower B, Hedley P, Machray G, Powell W (2004) A comparison of sequence-based polymorphism and haplotype content in transcribed and anonymous regions of the barley genome. Genome 47:389–398

Saal B, Wricke G (1999) Development of simple sequence repeat markers in rye (*Secale cereale* L.). Genome 42:964–972

Schneider K, Weisshaar B, Borchardt DC, Salamini F (2001) SNP frequency and allelic haplotype of *Beta vulgaris* expressed genes. Mol Breed 8:63–74

Senft P, Wricke G (1996) An extended genetic map of rye (*Secale cereale* L.). Plant Breed 115:508–510

Somers DJ, Kirkpatrick R, Moniwa M, Walsh A (2003) Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. Genome 46:431–437

Tenaillon MI, Sawkins MC, Long AD, Gaut B, Doebley JF, Brandon S (2001) Patterns of DNA sequence polymorphism

along chromosome 1 of maize (*Zea mays* ssp. *Mays* L.) Proc Natl Acad Sci USA 98:9161–9166

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development of cDNA derived microsatellite markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411–422

Thiel T, Kota R, Grosse I, Stein N, Graner A (2004) SNP2CAPS: a SNP and INDEL analysis tool for CAPS marker development. Nucleic Acids Res 32(1): e5

Thompson JD, Higgins DG, Gibson TJ (1994) Clustal-W —improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Tikhenko ND, Tsvetkova NV, Voylokov AV (2005) Genetic control of embryo lethality in crosses between common wheat and rye. Russ J Genet 41:877–884

Van K, Hwang E-Y, Young Kim M, Kim Y-H, Cho Y-I, Cregan PB, Lee S-H (2004) Discovery of single nucleotide polymorphisms in soybean using primers designed from ESTs. Euphytica 139:147–157

Varshney RK, Korzun V, Börner A (2004) Molecular marker maps in cereals: methodology and progress. In: Gupta PK, Varshney RK (eds) Cereal genomics. Kluwer Academic Publishers, The Netherlands, pp 35–82

Varshney RK, Sigmund R, Börner A, Korzun V, Stein N, Sorrells M, Langridge P, Graner A (2005) Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. Plant Sci 168:195–202

Zhu YL, Song QJ, Hyten DL, van Tassell C, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB (2003) Single-nucleotide polymorphisms in soybean. Genetics 163:1123–1134