Functional Molecular Markers in Barley: Development and Applications

R. K. Varshney[1], M. Prasad[1], R. Kota[1], R. Sigmund[1], A. Börner[1], J. Valkoun[2], U. Scholz[1], N. Stein[1] and A. Graner[1]

[1]Institute of Plant Genetics and Crop Plant Research (IPK), D-06466 Gatersleben, Germany;
[2]International Center for Agricultural Research in the Dry Areas (ICARDA), Aleppo, Syria,
e-mail: rajeev@ipk-gatersleben.de, r.k.varshney@cgiar.org

Abstract: The availability of sequence data from large-scale EST (expressed sequence tag) projects has made it possible to develop markers directly from genes. In order to develop functional markers and the preparation of a transcript map (integrating of genes/transcripts to genetic map) of barley, more than 1000 ESTs/cDNAs including 589 RFLP (Restriction Fragment Length Polymorphism), 255 SNP (Single Nucleotide Polymorphism) and 185 SSR (Simple Sequence Repeat or microsatellite) markers have been developed. These markers provide a good resource for a variety of purposes including pedigree analysis, the study of marker-trait associations and comparative mapping in other cereals. A computational study suggests a theoretical transferability of barley markers to wheat (95.2%), rice (70.3%), maize (69.3%), sorghum (65.9%), rye (38.1%) and even to dicot species (~16%). Comparative mapping of 9 barley EST-SSRs in rye suggested the potential of barley markers to saturate the genetic maps of minor crop species having a shortage of molecular markers on their genetic maps. A comparison of 50 SSR and 50 SNP marker data for cladistic analysis on a set of six diverse barley genotypes showed that both marker types yield similar groupings. Therefore, gene-derived markers are a good resource for sampling the functional diversity in natural germplasm collection or breeding material.

Keywords: gene-derived markers; functional markers; ESTs; transcript map; transferability; comparative mapping; functional diversity

Molecular markers are the most powerful diagnostic tools to detect DNA polymorphism both at the level of specific loci and at the whole genome level. In the past, these DNA-based molecular markers were developed either from genomic DNA libraries (RFLPs and microsatellites) or from random PCR amplification of genomic DNA (RAPDs- *R*andom *A*mplification of *P*olymorphic *D*NA*s*) or both (AFLPs- *A*mplified *F*ragment *L*ength *P*olymorphism*s*). These DNA markers can be generated in large numbers and can prove to be very useful for a variety of purposes relevant to crop improvement. For instance, these markers have been utilized extensively for the preparation of saturated molecular maps (genetical and physical). Their association with genes/QTLs controlling the traits of economic importance has also been utilized in some cases for indirect marker assisted selection (MAS). Other uses of molecular markers include gene introgression through backcrossing, germplasm characterization, genetic diagnostics, characterization of transformants, study of genome organization and phylogenetic analysis (Jain *et al.* 2002)

The availability of genomic DNA and cDNA sequences (ESTs) in public databases (e.g. http://www.ncbi.nlm.nih.gov; http://www.ebi.ac.uk) has made marker development more effective. The majority of the markers developed and used in the past belonged to genomic DNA, and therefore could belong to either the transcribed or the non-transcribed part of the genome without any information

available on their functions. In contrast, markers developed from coding sequences like ESTs or fully characterized genes frequently have been assigned known functions and therefore are known as functional markers (Andersen & Lubberstedt 2003). Functional markers can detect both length and sequence polymorphism in the expressed region of the genome. These markers are superior to random DNA markers (derived from genomic DNA) owing to complete linkage with potential trait locus alleles (Varshney *et al.* 2005a).

Keeping in view the importance of functional markers, we have exploited the barley ESTs generated at IPK in a systematic manner for genetic mapping. In this article, we describe the development of gene-derived markers and discuss the use of the developed resource for interspecific transferability and comparative mapping in cereals and diversity analysis in barley.

MATERIAL AND METHODS

Plant material and DNA isolation. To screen for markers, parental genotypes and DH (doubled haploid) lines of three mapping populations Igri, Franka, Steptoe, Morex and Oregon Wolfe Barley $OWB_{Dom} \times OWB_{Rec}$ along with Barke (reference genotype, since the majority of the barley ESTs at IPK were generated from Barke) were employed.

For diversity analysis, a set of six diverse barley (*Hordeum vulgare* L.) genotypes (obtained from ICARDA, Aleppo, Syria) namely 123722 (IG128088), 123793 (IG128159), 123804 (IG128170), 123807 (IG128173), 123834 (IG128200) and 123838 (IG128204) was used.

Total genomic DNA was extracted from leaf material as described in Graner *et al.* (1991).

RFLP analysis. A set of six restriction enzymes (*Bam*HI, *Hind*III, *Eco*RI, *Eco*RV, *Xba*I and *Dra*I) was used to digest genomic DNA. Southern blotting and probe labelling was carried out according to Graner *et al.* (1991). Autoradiography was performed by exposure of hybridised blots to imaging plates (Fuji Photo Film, Japan) and subsequent signal detection on a phosphoimager (Fuji, Japan).

SSR analysis. For development of EST derived SSR markers, 111 090 ESTs from the IPK Crop EST database (CR-EST; http://pgrc.ipk-gatersleben.de/cr-est/) were searched for all possible simple sequence repeats with the help of a PERL5 script. The method for SSR amplification and separation of SSR amplicons are described in Thiel *et al.* (2003).

For diversity analysis, microsatellite loci were amplified using fluorescent-dye labeled primers. Amplification products were separated on an ABI377 fragment analyzer and evaluated using GenoTyper 3.7 software (Applied Biosystems).

SNP analysis. To perform the PCR, EST-based primers were designed to amplify genomic DNA 7 genotypes mentioned above. All primers were designed using the Primer Express software (Perkin-Elmer, USA). To identify SNPs, PCR products amplified among the parental genotypes were sequenced in both directions on an ABI 377 automated sequencer using big dye-terminator chemistry (Perkin-Elmer, USA). DNA sequence data was checked for sequencing errors using the "Sequencher" software (Gene Codes Corporation, USA). Sequences were aligned using the GCG program pileup and polymorphisms between the parents were identified manually. The method for PCR and denaturing high-performance liquid chromatography was carried out on automated HPLC instrumentation equipped with a DNASep column as described in Kota *et al.* (2001a, b, 2003).

Preparation of consensus map. Polymorphic RFLP, SSR and SNP markers were mapped in respective mapping population. Finally, the consensus map was prepared by using JoinMap V3.0 using a LOD score of 3.0 (Stam 1993).

In silico ***transferability***. To examine theoretical transferability of barley EST-derived markers, BLASTN analysis (homology filter of expectation value ≤ 1.00E-10) of mapped barley SSR-ESTs was conducted against ESTs of 6 cereal (wheat, rye, rice, maize, sorghum and oats) and 3 dicot species (Arabidopsis, Medicago and Lotus), respectively. For this purpose, publicly available EST sequences for wheat, rye, rice, maize, sorghum, oats, Arabidopsis, Medicago and Lotus were acquired by the Sequence Retrieval System (SRS6, http://srs.ebi.ac.uk/) from the EMBL database [release 77 dt. 20 November 2003; query for: ORGANISM (*Triticum aestivum*/*Secale cereale*/*Oryza sativa*/*Zea mays*/*Sorghum*/*Avena*/*Arabidopsis*/*Medicago*/*Lotus*) and DIVISION (*EST*)]. The obtained 524 720 wheat, 9194 rye, 266 169 rice, 384 391 maize, 184 708 sorghum, 501 oats, 191 302 Arabidopsis, 188 642 Medicago and 36 311 Lotus ESTs were transferred into BLASTable databases by using the NCBI BLAST package ver. 2.26.

Diversity analysis. The 0/1 matrix obtained by selected SSR and SNP markers with the set of six genotypes was used for the calculation (NTSYSpc 2.1) of genetic dissimilarity according to Nei, SAHN

Table 1. Distribution of different type of markers on the transcript map of barley

Markers	Barley-chromosomes							Total
	1H	2H	3H	4H	5H	6H	7H	
RFLP	84	111	95	51	98	65	107	611
SNP	30	41	45	26	44	28	41	255
SSR	25	31	35	26	22	25	21	185
Total	139	183	175	103	164	118	169	**1051-12* (1039)**

*number of redundant ESTs mapped either as RFLPs, SSRs or SNPs

clustering and the construction of UPGMA (Unweighted Pair Group Method Arithmetic Average) phenograms. The PIC value, nucleotide diversity (π) for examined markers were computed as given in Thiel *et al.* (2003), Kota *et al.* (2003).

RESULTS AND DISCUSSION

Development of functional molecular markers

A flow chart for systematic development of gene-derived molecular markers is presented in Figure 1. In the context of a barley genomics programme, more than 111 090 ESTs from 22 different cDNA libraries representing different tissues at various stages were generated at IPK (Michalek *et al.* 2002; Zhang *et al.* 2004; http://pgrc.ipk-gatersleben.de/cr-est/). Cluster analysis, using stackPack software (Miller *et al.* 1999), yielded a tentative unigene set comprising of 25 000 genes. This unigene set was used for development of functional markers and construction of a transcript map of barley.

Using an RFLP assay, 543 cDNA clones detecting a total number of 558 loci were integrated into the genetic map of barley. Together with 46 cMWG markers detecting 53 cMWG loci that were earlier mapped by Graner *et al.* (1991) this amounts to a total number of 611 loci.

For development of SSR markers, the *MISA* (*MI*cro*SA*tellites, http://pgrc.ipk-gatersleben.de/misa) search module was used to screen the 111,090 barley ESTs that yielded 9564 redundant and 3122 non-redundant microsatellites (EST-SSRs). After analysing a total of 755 primer pairs in a set of 7 genotypes, 185 EST-SSR markers were developed.

Many of those ESTs, which could not be mapped as RFLPs or SSRs, were further analyzed for the presence of SNPs by allele-specific sequencing of EST-based amplicons. In addition, the SNiPpER alogrithm was used with 271 630 ESTs from different cDNA libraries, representing 23 barley varieties and 3069 candidate inter-varietal SNPs were identified (Kota *et al.* 2003). Using both of the above mentioned approaches a total of 255 non-

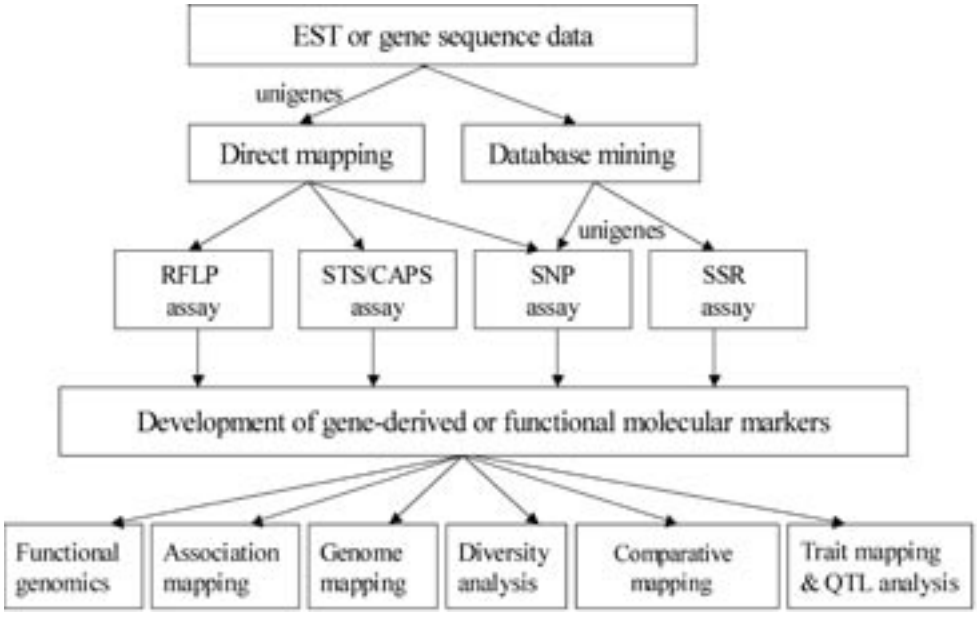

Figure 1. An overview on development and applications of gene-derived (functional) molecular markers

redundant SNP markers were mapped by DHPLC (Kota *et al.* 2001a, b).

Transcript map of barley

A consensus map was produced from the three individual maps by using JoinMap V3.0 programme. The highest number of markers was mapped in the OWB population followed by S × M and I × F. Altogether the transcript map contains a total of 1039 loci including 611 RFLP, 255 SNP and 185 SSR loci (Table 1). The map covers 1131.5 cM, with an average marker interval distance of 1.04 cM.

Applications of functional molecular markers

Interspecific transferability and comparative mapping. One of the most important features of gene-derived markers is their use in related species as gene sequences are known to remain highly conserved during evolution (Sorrells *et al.* 2003). BLASTN analysis of 974 mapped ESTs (from the transcript map) with the available 1 369 683 ESTs of six cereal species and 286 255 ESTs of three dicot species showed the presence of barley homologues in all the species, examined (Table 2). Among cereals, barley EST-derived markers showed *in silico* transferability as 95.2% in wheat, 70.3% in rice, 69.3% in maize and 65.9% in sorghum. A lower transferability of only 38.1%, which was observed for rye and 3.1% for oats has to be attributed to the small datasets (9194 ESTs in rye and 501 ESTs in oats), which were available for analysis and may be biased regarding the content of conserved sequences. Significant homology of barley ESTs with an average of 16% ESTs of even dicot species suggest that a conserved ortholog set (COS) of markers could be developed as demonstrated earlier by Fulton *et al.* (2002).

In fact, *in silico* comparative mapping (sequence comparison) of mapped barley SSR-ESTs against the mapping data of rye, wheat and rice indicated the presence of orthologues of the barley SSR-ESTs in the respective species (Varshney *et al.* 2005b). To confirm the *in silico* results, we mapped 9 barley EST-SSR markers in a mapping population of rye (P87 × P105) (Korzun *et al.* 2001). As a result, five loci were mapped on the homoeologous chromosome arms while four loci were mapped to non-homeologous chromosome arms in known translocations between the genomes of barley and rye, which therefore also represent syntenic linkage blocks. For instance, two barley markers namely GBM1008 (6HL) and GBM1046 (3HL) were mapeed on the rye linkage group 6 in the expected regions. These results suggest that developed EST-SSR markers can be used to increase the density of markers on the genetic maps of related plant species like rye, pearl millet where we do not have saturated maps (Varshney *et al.* 2005b; Wang *et al.* 2005). Thus, the developed functional markers will prove useful not only in comparative mapping among fairly divergent genomes but also for taxonomic studies and in deducing phylogenetic relationships between different genera and species (Gupta & Rustgi 2004).

Diversity analysis. To investigate the potential of gene-derived SSR and SNP markers for diver-

Table 2. Homology of mapped barley ESTs with ESTs of other plant species

Marker type	ESTs BLAST(N)ed	Wheat *Triticum aestivum*	Rye *Secale cereale*	Maize *Zea mays*	Sorghum *Sorghum bicolor*	Rice *Oryza sativa*	Oats *Avena sativa*	Arabidopsis *Arabidopsis thaliana*	Medicago *Medicago sativa*	Lotus *Lotus japonicus*
		(524,720 ESTs)	(9,194 ESTs)	(384,391 ESTs)	(184,708 ESTs)	(266,169 ESTs)	(501 ESTs)	(191,302 ESTs)	(188,642 ESTs)	(36,311 ESTs)
RFLP	536	511	210	388	372	393	25	107	111	87
SNP	253	243	99	191	174	186	6	47	41	36
SSR	185	173	62	96	96	106	3	18	16	12
Total	974	927 (95.2%)	371 (38.1%)	675 (69.3%)	642 (65.9%)	685 (70.3%)	34 (3.1%)	172 (17.7%)	168 (17.2%)	135 (13.9%)

Sequences of EST corresponding to mapped markers were BLAST(N)ed against the available ESTs of different plant species; Significant homology was considered if a barley EST gives a hit with the EST of a given species at an expectation value ≤ 1E-10

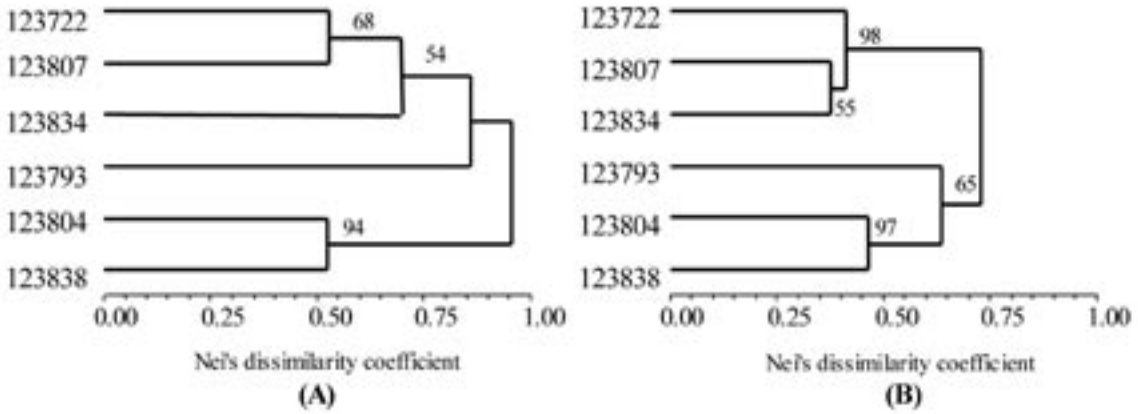

Figure 2. Comparison of SSR and SNP markers for diversity analysis

A total of 120 and 520 datapoints obtained by using 47 SSR and 46 SNP markers, respectively were used to prepare the phenograms shown as Figure 2A and Figure 2B, respectively. Bootstrap values (higher than 50) are given on nodes of branches in phenograms

sity anaylsis, a set of 50 SSR and 50 SNP markers randomly distributed throughout the genome (on all the linkage groups) were used with a set of six diverse barley genotypes.

A total of 47 out of 50 SSR markers were polymorphic in the examined genotypes. These markers yielded 2–4 (average 2.93) alleles with PIC values between 0.23 and 0.52 (average 0.36). As these SSR markers are derived from ESTs (conserved proportion of the genome), in comparison to genomic SSRs, they are characterized by lower number of alleles and lower PIC values (Pillen *et al.* 2000; Varshney *et al.* 2005a).

Similarly, 46 out of 50 EST-based SNP markers showed SNPs among the genotypes examined. Based on an average amplicon length of 409 bp (217–798 bp), 1 to 29 SNPs (average 7.6) were detected per EST-based SNP marker, yielding 2 to 6 haplotypes (average 4.1). The PIC values of identified SNPs and haplotypes ranged from 0.14–0.28 (average 0.23) and 0.44–0.83 (average 0.67), respectively. The calculated nucleotide diversity index (π) for each of the selected SNP marker was in the range of 0.16×10^{-2} to 2.79×10^{-2} with a mean 1.07×10^{-2}. The mean nucleotide diversity in examined genotypes is higher than that of the parental genotypes of the mapping populations (3.2×10^{-3}; our unpublished results), as the set of markers examined in this study is a pre-selected set of highly polymorphic markers.

In order to compare the results of SSR and SNP markers for cladistic analysis, a total of 120 and 520 data points obtained from 47 SSR and 46 SNP markers, respectively were used to construct phenograms for each marker type. Both phenograms classify the examined genotypes in a similar way. In both phenograms, three genotypes (123722, 123807 and 123834) group in one cluster, two genotypes (123804 and 123838) in another cluster while one genotype (123793) is very distant to the above clusters (Figure 2). These results are in accordance with earlier studies where it was concluded that different types of markers derived from genes are more suitable for fingerprinting studies as they yield similar groupings of unrelated germplasms (Kota *et al.* 2001a) and useful resources for assessment of functional diversity (Eujayl *et al.* 2001).

CONCLUSIONS

A comprehensive resource for gene-derived markers including SSRs and SNPs has been developed for barley. The widespread application of these functional markers in a variety of studies will have far reaching implications relevant to crop improvement. The transcript map enables barley geneticists to rapidly identify target regions in rice, for an efficient marker saturation of defined regions in the barley genome. The present study shows the examples of applications of functional molecular markers in interspecific transferability and comparative mapping in cereal species and diversity analysis in barley.

Acknowledgements. We thank C. Künzel, U. Beier, J. Pohl, S. Stegmann, M. Kretschmann and J. Perovic for their technical assistance. Thanks are also due to H. Miehe

for his help in bioinformatic analysis. This work was supported by grants from the Federal Ministry of Education and Research (BMBF), the State of Saxony-Anhalt (LSA) and the Grains Research and Development Corporation (GRDC) Australia.

References

Andersen J.R., Lubberstedt T. (2003): Functional markers in plants. Trends in Plant Science, **8**: 554–560.

Eujayl I., Sorrells M., Baum M., Wolters P., Powell W. (2001): Assesment of genotypic variation among cultivated durum wheat based on EST-SSRs and genomic SSRs. Euphytica, **119**: 39–43.

Fulton T.M., van der Hoeven R., Eannetta N.T., Tanksley S.D. (2002): Identification, analysis and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell, **14**: 1457–1467.

Graner A., Jahoor A., Schondelmaier J., Siedler H., Pillen K., Fischbeck G., Wenzel G., Hermann R.G. (1991): Construction of an RFLP map of barley. Theoretical and Applied Genetics, **83**: 250–256.

Gupta P.K., Rustgi S. (2004): Molecular markers from the transcribed/expressed region of the genome in higher plants. Functional & Integrative Genomics, **4**: 139–162.

Jain S.M., Brar D.S., Ahloowalia B.S. (2002): Molecular Techniques in Crop Improvement. Kluwer Academic Publishers, Dordrecht.

Korzun V., Malyshev S., Voylokov A.V., Borner A. (2001): A genetic map of rye (*Secale cereale* L.) combining RFLP, isozyme, protein, microsatellite and gene loci. Theoretical and Applied Genetics, **102**: 709–717.

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

Kota R., Varshney R.K., Thiel T., Dehmer K.J., Graner A. (2001a): Generation and comparison of EST-derived SSRs and SNPs in barley (*Hordeum vulgare* L.). Hereditas, **135**: 145–151.

Kota R., Wolf M., Michalek W., Graner A. (2001b): Application of denaturing high-performance liquid chromatography for mapping of single nucleotide polymorphisms in barley (*Hordeum vulgare* L.). Genome, **44**: 523–528.

Michalek W., Weschke W., Pleissner K.P., Graner A. (2002): EST analysis in barley defines a unique set comprising 4000 genes. Theoretical and Applied Genetics, **104**: 97–103.

Miller R.T., Christoffels A.G., Gopalakrishnan C., Burke J., Ptitsyn A.A., Broveak T.R., Hide W.A. (1999): A comprehensive approach to clustering of expressed human gene sequence: the sequence tag alignment and consensus knowledge base. Genome Research, **9**: 101–113.

Pillen K., Binder A., Kreuzam B., Ramsay L., Waugh R., Forster J., Leon J. (2000): Mapping new EMBL-derived barley microsatellites and their use in differentiating German barley cultivars. Theoretical and Applied Genetics, **101**: 652–660.

Sorrells M.E., Rota M.L., Bermudez-Kandianis C.E., Greene R.A., Kantety R., Munkvold J.D., Miftahudin M., Mahmoud A., Ma X., Gustafson P.J., *et al.* (2003): Comparative DNA sequence analysis of wheat and rice genomes. Genome Research, **13**: 1818–1827.

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

Thiel T., Michalek W., Varshney R.K., Graner A. (2003): Exploiting EST database for the development and characterization of gene derived SSR-markers in barley (*Hordeum vulgare* L.). Theoretical and Applied Genetics, **106**: 411–422.

Varshney R.K., Graner A., Sorrells M.E. (2005a): Genic microsatellite markers in plants: Features and applications. Trends in Biotechnology, **23**: 48–55.

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

Wang M.L., Barkley N.A., Yu J.-K., Dean R.E., Newman M.L., Sorrells M.E., Pederson G.A. (2005): Transfer of simple sequence repeat (SSR) markers from major cereal crops to minor grass species for germplasm characterization and evaluation. Plant Genetics Resources: Characterization and Utilization, **3**: 45–58.

Zhang H., Sreenivasulu N., Weschke W., Stein N., Rudd S., Radchuk V., Potokina E., Scholz U., Schweizer P., Zierold U., Langridge P., Varshney R. K., Wobus U., Graner A. (2004): Large-scale analysis of the barley transcriptome based on expressed sequence tags. Plant Journal, **40**: 276–290.