P L A N T  B R E E D I N G  A N D  S E E D  S C I E N C E
Volume 48 (no. 2/2) 2003

B. Hackauf, P. Wehling

Federal Centre for Breeding Research on Cultivated Plants, Institute of Agricultural Crops, Rudolf–Schick–Platz 3a, D–18190 Gross Luesewitz, Germany

# DEVELOPMENT OF MICROSATELLITE MARKERS IN RYE: MAP CONSTRUCTION

## ABSTRACT

Rye ESTs from public sequence databases proved to be a valuable resource to develop microsatellite markers in rye. One hundred and twenty–one EST–derived *Secale cereale* microsatellites (SCM) were genetically analyzed in a BC1 population. Fourty–three percent of the studied SCM markers displayed a polymorphism in this mapping population. Linkage analysis in relation to genomic SSR anchor markers as well as AFLP markers allowed to arrange 41 polymorphic EST–derived SCM markers into seven linkage groups which correspond to the seven rye chromosomes. In total, 56 rye SSR markers could be integrated in this second–generation linkage map of rye comprising 685 cM of the rye genome. Distorted segregations with excess of heterozygous progeny were found for most of the markers on chromosome 7R and possible reasons for these are discussed.

*Key words:* distorted segregation, ESTs, functional map microsatellites, *Secale cereale* L.

## INTRODUCTION

The genetic linkage map of rye (*Secale cereale* L.) has continuously been developed during the past two decades initially by use of isozyme markers (Wehling, 1985) and later on by including different types of DNA markers (Devos *et al.* 1993; Senft and Wricke, 1996; Korzun *et al.*, 1998, 2001; Ma *et al.*, 2001; Masojć *et al.*, 2001). Among the latters, microsatellites or simple–sequence repeats (SSRs) appear as the most attractive marker class. SSRs are defined as short, tandemly arrayed sequence motifs of usually two, three or four base pair repeat units. The number of repeat units is highly variable leading to multiple allelism at individual SSR loci. Use of sequence–specific primers located in the flanking regions of a microsatellite allows SSRs to be easily tagged by means of PCR. Further advantages of microsatellites for application in genetics and molecular breeding are their codominant inheritance, rel–

*Communicated by Piotr Masojć*

ative abundance, extensive genome coverage and low requirement of sample DNA (Powell *et al.* 1996).

Rye microsatellites were first described by Saal and Wricke (1999), and twelve of them were mapped by these authors on chromosomes *1R*, *2R*, *3R*, *5R*, *6R* and *7R*. The number of rye microsatellite markers, though, has to be significantly increased to be of practical significance. Ideally, construction of a linkage map consisting of evenly distributed SSRs which cover the entire rye genome would provide a valuable tool for rye as well as triticale genetists and breeders.

Sequence information for the development of SSR markers may be obtained by establishing subgenomic libraries and subsequent screening for clones bearing SSRs, which is an expensive, laborious and time–consuming task. As an alternative in the genomics era, an increasing number of sequences deposited in public sequence databases may serve as a data mine for SSR–marker development. Recently deposited expressed–sequence tag (EST) data from different rye tissues served a valuable resource for the development of 157 novel *Secale cereale* microsatellite (SCM) markers (Hackauf and Wehling, 2001). In the present paper, mapping of a subset of these SCM markers in the rye genome is reported.

## MATERIAL AND METHODS

### Plant materials:

The BC1 population 9953 was kindly provided by H. Wortmann (Hybro GmbH & Co. KG, Bad Schönborn, Germany) for segregation and linkage analyses. This population originated from a cross of an inbred line and a non–adapted genebank accession. A single $F_1$ plant had been crossed to the BC1 with the inbred line as pollen parent (Miedaner, pers. comm.). DNA from individual plants was extracted from two–week old leaves using standard methods.

### SSR–marker analysis:

The majority of microsatellites investigated in this study were derived from rye ESTs as described elsewhere (Hackauf and Wehling, 2001). Genomic microsatellite–anchor markers were used for the chromosomal localisation of linkage groups (Table 1). Information on the chromosomal assignments of these markers was published by Saal and Wricke (1999) or kindly communicated by V. Korzun (pers. comm.), respectively. Anchor marker *Xscm1* was developed from a rye pollen–cDNA fragment encoding a thioredoxin–h, which was previously mapped on chromosome 1RS (Hackauf, 1999). SSR loci are designated according to the recommended rules for gene symbolization in wheat and related species (http://wheat.pw.usda.gov/ggpages/maps.html).

PCR components were combined using the MultiprobeII Liquid Handling System (Canberra–Packard). SSRs were assayed on LI–COR

Model 4200 automated fluorescent DNA sequencers (LI–COR Inc., Lincoln, NE) as described previously (Hackauf and Wehling, 2001).

Table 1

**Rye–SSR anchor markers.**

| Locus[a] | Reference |
|---|---|
| *Xscm1–1R* | Hackauf, 1999 |
| *Xscm5–3R*[b] | Saal and Wricke, 1999 |
| *Xscm43–2R*[b] | Saal and Wricke, 1999 |
| *Xscm138–5R*[b] | Saal and Wricke, 1999 |
| *Xscm180–6R*[b] | Saal and Wricke, 1999 |
| *Xlprm2–4R* | Korzun, pers. comm. |
| *Xlprm4–7R* | Korzun, pers. comm. |
| *Xlprm7–7R* | Korzun, pers. comm. |
| *Xlprm10–2R* | Korzun, pers. comm. |
| *Xlprm11–5R* | Korzun, pers. comm. |
| *Xlprm15–1R* | Korzun, pers. comm. |
| *Xlprm19–4R* | Korzun, pers. comm. |

[a] Loci are given according to the recommended rules for gene symbolization in wheat and related species (http://wheat.pw.usda.gov/ggpages/maps.html).
[b] These markers were published as SCM5, SCM43, SCM138 and SCM180, respectively

**Segregation analysis:**

SSR–segregation data was collected for 88 out of 109 individuals of the BC1–mapping population and analyzed using the software package JoinMap v.3.0 (Van Ooijen and Voorrips, 2001). The Kosambi function was used to convert recombination values to genetic distances (cM). Linkage groups were separated using a LOD score of 5.0. Data on AFLP segregation among 90 individuals and arrangements of AFLP markers into seven major and 3 minor anonymous linkage groups was kindly provided by H.H. Geiger and T. Miedaner (Univ. of Hohenheim, State Plant Breeding Institute, Hohenheim, Germany).

## RESULTS

A total of 121 microsatellites were tested which had been derived from ESTs expressed in anthers, cold–stressed seedlings, or $Al^{3+}$–stressed and unstressed roots of rye, respectively. Fifty–two (43%) of these SSRs proved to be polymorphic in the BC1–mapping population. Thirty–eight markers (73%) displayed a codominant inheritance with a segregation pattern as expected for this population type (Fig.1), while a dominant/recessive inheritance could be observed for 14 SCM markers. Linkage analysis in relation to SSR–anchor markers allowed to arrange 41 polymorphic EST–derived SCM markers into a total of seven linkage

groups which correspond to the seven rye chromosomes (Fig. 2). Eleven polymorphic SCM markers could not be integrated into the linkage map.

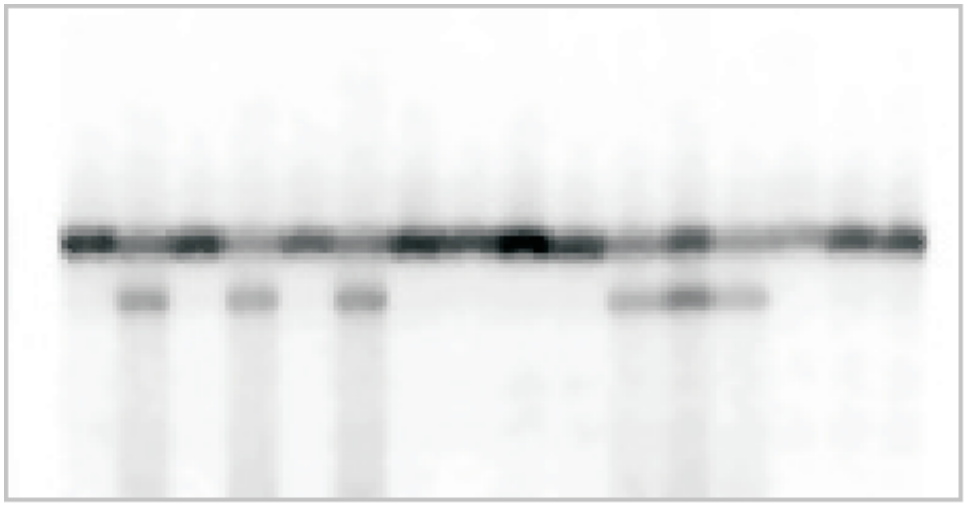

Fig. 1 Codominant segregation patern of the EST–derived rye microsatelite marker *Xscm77* in the $BC_1$ population9953

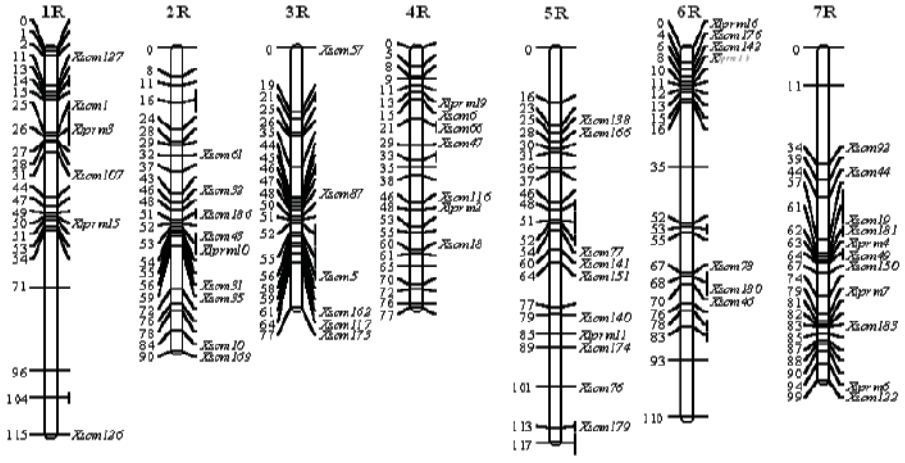


Fig. 2 A second–generation linkage map of rye based on SSRs and AFLPs

Genomic distribution of EST–derived SCM markers appeared to be uniform, although a number of closely linked markers were present in the distal part of chromosome 6RS as well as in the centromeric part of chromosome 7R. In general, marker segregation was according to expectation on six of the seven chromosomes. There was a clear contrast, however, with chromosome 7R where distorted segregations were found for 23 out of 28 markers, with 19 markers deviating significantly from expectation (not shown). All but one distorted segregations were characterized by excess of heterozygotes. Among the markers on chromosome 7R all eleven SSR markers were affected, seven of them significantly (Table 2). For *Xlprm4*, which displayed the most pronounced deviation from the expected 1:1 ratio the observed frequency, $f_M$, of the recurrent homozygous marker genotype equaled 0.3. Markers flanking *Xlprm4* on both sides displayed increased $f_M$ values which corresponded to increasing recombination frequencies, *r*, between the respective markers and *Xlprm4*.

Table 2

**Distorted segregation of SSR markers on chromosome 7R**

| Locus [a] | a | h | $\chi^2$ | Df | Signif. [b] | $f_M$ *obs.* [c] | $f_M$ *exp.* [d] | r [e] |
|---|---|---|---|---|---|---|---|---|
| *Xscm92–7R* | 39 | 46 | 0.6 | 1 | – | 0.459 | 0.39 | 0.227 |
| *Xscm44–7R* | 32 | 45 | 2.2 | 1 | – | 0.416 | 0.38 | 0.209 |
| *Xscm181–7R* | 21 | 38 | 4.9 | 1 | ** | 0.356 | 0.31 | 0.020 |
| *Xscm19–7R* | 28 | 56 | 9.3 | 1 | **** | 0.333 | 0.31 | 0.013 |
| *Xlprm4–7R* | 23 | 54 | 12.5 | 1 | ****** | 0.299 | 0.30 | – |
| *Xscm49–7R* | 14 | 23 | 2.2 | 1 | – | 0.378 | 0.30 | 0 |
| *Xscm150–7R* | 28 | 54 | 8.2 | 1 | **** | 0.341 | 0.32 | 0.041 |
| *Xlprm7–7R* | 25 | 40 | 3.5 | 1 | * | 0.385 | 0.39 | 0.231 |
| *Xlprm6–7R* | 28 | 49 | 5.7 | 1 | ** | 0.364 | 0.41 | 0.273 |
| *Xscm122–7R* | 33 | 50 | 3.5 | 1 | * | 0.398 | 0.42 | 0.311 |
| *Xscm183–7R* | 29 | 37 | 1 | 1 | – | 0.439 | 0.43 | 0.321 |

[a] Order of markers is according to their recombination with *Xlprm4*
[b] Significance levels: *:0.1, **0.05, ***:0.01, ****0.005, *****0.001, ******0.0005
[c] $f_M = a/(a+h)$
[d] for $z = 0.43$ and $r = 0$ (see discussion)
[e] *r* is the recombination frequency between with *Xlprm4*

Compared to the underlying AFLP–framework map inclusion of the 56 EST–derived as well as genomic rye SSR markers resulted in an extension of the total length of the BC1–linkage map by approximately 120 cM. As a result, a linkage map comprising 685 cM was constructed which is solely based on PCR markers. Since EST–derived SSR markers represent expressed genes they allow to establish “functional maps”. Fifteen of the mapped EST–derived SSR markers could be associated with proteins of known or unknown function (Table 3).

Table 3

**Mapped rye ESTs with significant similiarity (BlastX) to known genes**

| Locus | Function [a] |
|---|---|
| *Xscm1–1R* | thioredoxin–like protein |
| *Xscm10–2R* | farnesylated protein (ATFP6) |
| *Xscm47–4R* | putative MYB family transcription factor |
| *Xscm66–4R* | ribonucleotide reductase R2 |
| *Xscm87–3R* | hypothetical protein |
| *Xscm92–7R* | isomerase like protein |
| *Xscm107–1R* | probable DNA–binding protein GBP16 |
| *Xscm116–4R* | acyl–CoA–binding protein |
| *Xscm117–3R* | cysteine proteinase inhibitor |
| *Xscm150–7R* | myo–inositol 1–phosphate synthase |
| *Xscm151–5R* | putative UDP–glucose dehydrogenase |
| *Xscm166–5R* | apospory–associated protein C–like |
| *Xscm169–2R* | proline–rich protein |
| *Xscm179–5R* | putative cellulase |
| *Xscm186–2R* | cold–responsive protein COR14a |

[a] Details on the BLASTX similarity are described elsewhere (Hackauf and Wehling, 2001)

## DISCUSSION

Results presented allow to include rye in the list of agronomically important grass species for which linkage maps based on significant numbers of genomic or EST–derived SSR markers are available. From SSR maps of other grass species such as barley (Liu *et al.*, 1996, Pillen et al, 2000), maize (Senior *et al.* 1996), rice (Temnykh *et al.* 2000) or wheat (Röder *et al.*, 1998) it is known that microsatellite markers generally display a uniform distribution over the genome. The BC1–mapping data on rye reported in the present paper are in agreement with this. Covering all the seven chromosomes, the SCM markers mapped so far provide a good starting–point for further extending this second–generation linkage map of rye and should serve valuable anchor markers for mapping purposes in other experimental populations.

Regular segregation ratios could be observed for most SCM markers except of markers on chromosome 7R which partly showed significantly distorted segregations. Even 7R markers deviating not significantly from the 1:1 expectation displayed ratios which were similarly skewed towards the heterozygotes. This is also true for marker *Xscm49* which cosegregates with *Xlprm4* but is not significant in deviation due to the low number of $BC_1$ individuals scored. Deviations from expected Mendelian segregation patterns were also reported by other authors for mo–

lecular markers in almost all rye–mapping populations analyzed and were observed for markers on chromosomes 1R, 2R, 4R, 5R, 6R, and 7R (Philipp *et al.*, 1994; Wanous *et al.*, 1995; Senft and Wricke, 1996; Korzun *et al.*, 1998, 2001; Ma *et al.*, 2001). One possible explanation for the distorted BC1 segregation data presented above is the presence of a locus governing zygotic selection on chromosome *7R*. In this hypothet–ical case, the most probable location of the selective locus would be nearby or at the *Xlprm4* marker locus. This marker displays the most pronounced excess of heterozygous over homozygous progeny among all SSR markers on chromosome *7R*, as reflected by its observed $f_M$ value, and is flanked by markers *Xscm19* and *Xscm150* which map only 1.2 cM and 4.3 cM apart from *Xlprm4*, respectively. Under selection the ex–pected fraction, $f_M$, of progeny which is homozygous for the recur–rent–parent allele at a marker locus *M* depends on two parameters, namely the zygotic selection coefficient, $z$, and the recombination $r$ be–tween the marker and the selectively active locus, with $f_M = (z(1-r)+r)/(1+z)$ and $0 = z = 1$ and $0 = r = 0.5$ (for details, see Weber and Wricke, 1994). If we assume that the selective locus is tightly linked with marker *Xlprm4* then in an approximation, $r$ can be set zero for the recombination between *Xlprm4* and the hypothetical, selective locus. For this case, the observed frequency $f_M = 0.299$ of progeny homozygous at the *Xlprm4* locus may be used to estimate $z \approx 0.43$, i.e., zygotic selec–tion at the hypothetical locus would not be absolute but of medium in–tensity. For the assumption of $z = 0.43$ the expected $f_M$ values may be calculated for all SSR markers on chromosome *7R* on the basis of their recombination with *Xlprm4*, the latter of which is assumed to be tightly linked with negligible recombination to the selective locus. Table 2 shows that under these assumptions, the observed $f_M$ values for the eleven SSR markers are in good agreement with the expected ones.

For the alternative hypothesis, i.e., gametic selection on the female side practically the same figures for expected $f_M$ values at marker loci are obtained (not shown) if a gametic selection coefficient $g \approx 0.30$ is as–sumed. This value can be deduced from $f_M = g(1-2r)+r$ with $0 = g = 0.5$, using the same empirical settings for $f_M$ and $r$ as given above. Whether selection of zygotes or female gametes is responsible for the disturbed segregations of chromosome *7R* markers cannot be decided from the present data.

In their mapping study Senft and Wricke (1996) solely observed zy–gotic selection which occured at eight marker loci in an F2–mapping population according to chi–square analysis. One of these markers, *Xiag111b* was located on chromosome 7R and displayed symmetric defi–ciencies of both homozygous marker classes (Senft, 1996). In the $BC_1$ population of the present study zygotic selection may have occurred against genotypes homozygous for the recurrent parent allele or in fa–vour of the heterozygous state. Zygotic selection most probably would have affected seedset or germination.

EST markers are attractive due to their "expressed" nature which opens up the possibility to construct molecular linkage maps representing gene–rich regions of the genome. Once EST markers have been derived from a major fraction of the expressed genes in an organism, high–density EST maps may be constructed which not only allow to quickly identify trait gene markers but also to directly tag and isolate a gene of interest. Thus, EST mapping is a valuable tool in genomics. Mapping of ESTs still relies mainly on well known molecular marker systems like RFLP (Harushima *et al.*, 1998), SSCP (Schneider *et al.* 1999) or DGGE (Temesgen *et al.* 2001). As a consequence, EST mapping remains a relatively time–consuming and laborious venture although novel approaches were proposed recently (Cato *et al.* 2001). In addition, EST markers monitored by means of the above–mentioned methods generally display a low degree of polymorphism which may limit their use in breeding programmes. As an alternative, ESTs containing simple–sequence repeats may be used to develop easy–to–use SSR markers for genetic and breeding purposes. EST–derived microsatellites combine the advantages of representing expressed genes, displaying high polymorphisms and being based solely on PCR–based methodology without further processing. Furthermore, in highly repetitive genomes EST–derived SSR markers allow for higher–quality readings as compared to genomic SSRs in that stuttering and multiple loci are reduced (Scotti *et al.*, 2000; Hackauf and Wehling, 2001).

## REFERENCES

Cato, S.A., Gardner R. C., Kent J., Richardson T. E. 2001. A rapid PCR–based method for genetically mapping ESTs. Theor. Appl. Genet. 102, 296–306.

Devos, K. M., Atkinson M. D., Chinoy C. N., Francis H. A., Harcourt R. L., Koebner R. M. D., Liu C. J., Masojć P., Xie D. X., Gale M. D. 1993. Chromosomal rearrangements in the rye genome relative to that of wheat. Theor. Appl. Genet. 85, 673–680.

Hackauf, B. 1999. Molekulare Charakterisierung des Selbstinkompatibilitätslocus *S* bei Roggen (*Secale cereale* L.). Ph.D. thesis, University of Hannover, Germany.

Hackauf B., Wehling P. 2001. Identification of microsatellite polymorphisms in an expressed portion of the rye genome. Plant Breed. 121, 17–25.

Harushima Y., Yano M., Shomura A., Sato M., Shimano T., Kuboki Y., Yamamoto T., Lin S. Y., Antonio B. A., Parco A., Kajia H., Huang N., Yamamoto K., Nagamura Y., Kurata N., Khush G. S., Sasaki T. 1998. A high–density rice genetic linkage map with 2275 markers using a single F2 population. Genetics, 479–494.

Korzun V., Malyshev S., Kartel N., Westermann T., Weber W. E., Börner A. 1998. A genetic linkage map of rye (*Secale cereale* L.). Theor. Appl. Genet. 96, 203–208.

Korzun V., Malyshev S., Voylokov A. V., 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.

Liu, Z. W., Biyashev R. M., Saghai Maroof M. A. 1996. Development of simple sequence repeat markers and their integration into a barley linkage map. Theor. Appl. Genet. 93, 869–876.

Ma, X.–F., Wanous M. K., Houchins K., Rodriguez M. A., Milla P. Goicoechea G., Wang Z., Xie M., Gustafson., J. P. 2001. Molecular linkage mapping in rye (*Secale cereale* L.). Theor. Appl. Genet. 102, 517–523.

Masojć P., Myśków B., Milczarski P. 2001. Extending a RFLP–based genetic map of rye using random amplified polymorphic DNA (RAPD) and isozyme markers. Theor. Appl. Genet. 102, 1273–1279.

Ooijen J.W. van, Voorrips R.E. 2001. JoinMap® Version 3.0, Software for the calculation of genetic linkage maps. Plant Research International, Wageningen, the Netherlands

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 bar–ley cultivars. Theor. Appl. Genet. 101, 652–660.
Philipp U., Wehling P., Wricke G. 1994. A linkage map of rye. Theor. Appl. Genet. 88, 243–248.
Powell W., Machray G. C., Provan J. 1996. Polymorphism revealed by simple sequence re–peats. Trends Plant Sci. 1, 215–222.
Röder M. S., Korzun V., Wendehake K., Plaschke J.,. Tixier M.–H, Leroy P., Ganal M. W., 1998: A microsatellite map of wheat. Genetics 149, 2007–2023.
Saal B., Wricke G. 1999: Development of simple sequence repeat markers in rye (*Secale cereale* L.). Genome 42, 964–972.
Schneider, K., D. Borchardt C., Schäfer–Pregl R., Nagl N., Glass C., Jeppsson A, Gebhardt C., Salamini F. 1999. PCR–based cloning and segregation analysis of functional gene hoologues in *Beta vulgaris*. Mol. Gen. Genet. 262, 515–524.
Scotti I., Magni F., Fink R., Powell W., Binelli G., Hedley P.E. 2000: Microsatellite repeats are not randomly distributed within Norway spruce (*Picea abies* K.) expressed sequences. Genome 43, 41–46.
Senft, P. (1996) Erweiterung der genetischen Kopplungskarte und Markierung wirtschaftlich wichtiger Gene bei Roggen (*Secale cereale*. L.). PhD thesis, University of Hannover, Germany.
Senft P., Wricke G. 1996. An extended genetic map of rye (*Secale cereale* L.). Plant Breed. 115, 508–510.
Senior M. L., Chin E. C. L., Smith J. S. C., Stuber C. W. 1996. Simple sequence repeat mark–ers developed from maize sequences found in the GenBank database: Map construction. Crop Sci. 36, 1676–1683.
Temesgen B., Brown G. R., Harry D. E., Kinlaw C. S., Sewell M. M., Neale D. B. 2001. Genetic mapping of expressed sequence tag polymorphism (ESTP) markers in loblolly pine (*Pinus taeda* L.). Theor. Appl. Genet. 102, 664–675.
Temnykh S., Park W. D., Ayres N., Cartinhour S., Hauck N., Lipovich L., Cho Y. G., Ishii T., McCouch S. R. 2000. Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.) Theor. Appl. Genet. 100, 697–712.
Wanous M. Goicoechea K. P., Gustafson J. P. 1995. RFLP maps of rye chromosomes 6R and 7R including terminal C–bands. Genome 38, 999–1004.
Weber W. E., Wricke G. 1994. Genetic Markers in Plant Breeding. Adv. Plant Breed. 16. Paul Parey, Berlin, Hamburg.
Wehling P. (1985) Electrophoretic analyis of 10 enzyme systems in rye: linkage relationships and chromosomal location of isozyme loci. In: Proc. EUCARPIA Meeting of the Cereal Section on Rye. Svalöf, Sweden, pp101–124.