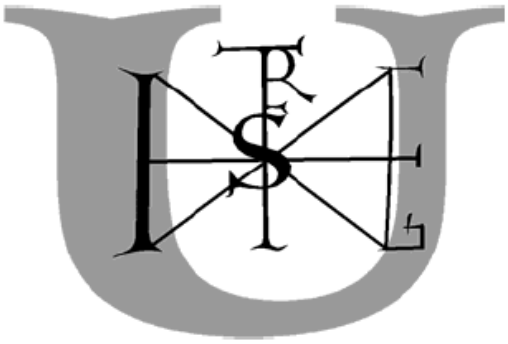

Emese Nagy:
Molecular polymorphisms of maize inbred lines (Zea mays)

PhD Thesis
SzIE, Gödöllő, 2004

Advisors

Cs. L. Marton

G Gyulai

Ph.D. School:	Plant Sciences
Scientific Code:	4. Agricultural Sciences
	4.1. Agronomy and Horticulture
Ph.D. Program:	Plant breeding, genetics and biotechnology
Director:	Prof. Dr. Ferenc Virányi, Ph.D., DSc
	SZIU Department of Plant Protection
Scientific Secretary:	Dr. Gábor Gyulai, Ph.D., CSc
	SZIU Department of Genetics and Plant Breeding

INTRODUCTION

As the result of rapid developments, molecular biological and genetic methods are becoming more and more widely used in various fields of plant breeding, including maize breeding. In addition to DUS (UPOV) traits based on phenotypic plant characteristics, the use of molecular markers allows the genetic stock to be analysed for genetic variability and genetic distances on the basis of polymorphisms between breeding lines, so that variety descriptions can be compiled for identification and variety protection purposes (*Smith and Senior,* 1999).

The grouping of maize varieties according to genetic relationships also plays an important role in breeding, since one of the basic preconditions for heterosis breeding is a sufficiently great genetic distance between the parental lines. The more precise the description of the varieties, allowing even fine differences between the genotypes to be detected, the more successfully they can be divided into related groups. Molecular techniques which, directly or indirectly, map the genetic background of the plants are indispensable for this work (*Pejic et al.*, 1998).

AIMS

The aims of the research were the following:

1. The analysis of polymorphism in 46 maize inbred lines bred in Martonvásár or used as parental lines in Martonvásár hybrids, on the basis of morphological descriptions, isoenzyme patterns and analysis using DNA-based methods: RAPD and gene-linked microsatellite (SSR) markers.
2. Pedigree analysis on the 46 maize inbred lines, tracing the origin of the lines back to the initial populations from which they were derived.
3. Determination of the allele frequency of the isoenzyme loci in 294 maize inbred lines used as parental lines in Martonvásár hybrids and in 119 other inbred lines obtained in the course of international cooperation; comparison of the results with data from the literature.
4. Elaboration of a system allowing data from morphological descriptions to be compared with the PIC (polymorphic index content) values used to evaluate the results of polymorphism analysis on genetic markers.
5. Classification of maize inbred lines in related groups in terms of morphological, biochemical and genetic data, on the basis of data from polymorphism analysis.
6. Investigation of correlations between pedigree data and morphological, biochemical and genetic markers using regression analysis.
7. Investigation of how the location on the dendrogram of lines not belonging to any of the major related groups compares with observations made by breeders.
8. The final aim was to elaborate a system consisting of both morphological descriptions and of the optimum number of biochemical and genetic markers or marker combinations required to give a realistic picture of relationships between inbred lines, confirmed by the results of pedigree analysis. This would allow inbred lines of unknown origin to be classified in related groups, making the system suitable for use by breeders when planning crossing programmes.

MATERIALS AND METHODS

Plant material

The work was carried out on 46 maize inbred lines with known genetic backgrounds, chosen from the major related groups: Lancaster, Iodent, Iowa Stiff Stalk Synthetic (ISSS), Mindszentpusztai Yellow Dent (MYD) and OP Lacaune. In addition, use was made of a number of lines which cannot be classified in any of these major groups, but whose origin is precisely known: Argentinian flints, early Canadian lines, lines related to *W 117* and derivatives of *Co 125*.

The software available from the *International Crop Information System* (ICIS) was used for pedigree analysis. As a first step the origin of the lines was traced back to the initial populations, as suggested by *Gerdes et al.* (1994). Coefficients of relationship were then calculated according to the program by comparing pairs of lines.

Polymorphism analysis

Polymorphisms and degrees of relationship were determined based on morphological descriptions, isoenzyme patterns and analysis at the DNA level, involving RAPD and gene-linked microsatellite (SSR) markers.

On the basis of the isoenzyme patterns, the allele frequency of enzyme loci was determined for 294 maize inbred lines bred in Martonvásár and for a further 119 lines obtained through international cooperation. The results were compared with those reported by *Goodman and Stuber* (1983), who carried out allele frequency studies on the isoenzyme loci of 250 inbred maize lines native to South America.

Traits scored for the morphological descriptions

The inbred lines were sown at two locations in Hungary (Martonvásár and Mezőkövesd) in separate rows in two-year, small-plot trials, arranged in a random block design with four replications. The traits were scored on the average of 10 plants per plot according to the guidelines provided by UPOV TG 2/6 (1994) and *Smith and Smith* (1989a,b).

The genetic distance between the inbred lines, the origin of which was traced back to the initial populations according to *Gerdes et al.* (1994), was determined by comparing pairs of lines using the V1.0 genetic data analysis software provided by the *International Crop Information System* (ICIS).

Isoenzyme analysis

The isoenzymes were identified by means of starch gel electrophoresis, using the method of *Goodman and Stuber (1983)* and *Stuber et al.* (1988).

Five seeds of each maize inbred line were analysed for the isoenzymes laid down in the UPOV TG 2/6 regulations, namely: maleic acid dehydrogenase (6 loci), isocitrate dehydrogenase (2 loci), 6-phosphogluconate dehydrogenase (2 loci), phosphoglucomutase (2 loci), phosphoglucose isomerase (1 locus), acidic phosphatase (1 locus) and alcohol dehydrogenase (1 locus).

Analyses involving DNA markers

In the course of the genetic analyses the 46 maize inbred lines were tested with 40 primers belonging to two major primer groups (RAPD, SSR). The 10 bp OP/AB set (1–20) manufactured by Operon Sci. (USA) was used to analyse random polymorphisms, while genetic polymorphisms occurring in repetitive DNA sequences were analysed using SSR primer pairs with a length of 16–18 bp (*Weining and Langridge* 1991).

Statistical methods used to evaluate the analytical data

In the case of morphological descriptions the measured morphological data were analysed using two-factor analysis of variance. For the polymorphism analysis the differences between the lines were estimated on the basis of the UPOV TG 2/6 application form, applying the rules laid down by UPOV and converting the measured data into scores.

In the polymorphism analysis Dice similarity indexes were calculated by comparing pairs of lines, and the values of PIC (polymorphic index content) were determined, expressing the discriminating ability at any given locus.

The degrees of relationship were analysed by means of hierarchical cluster analysis.

For the purposes of linear regression analysis, the lines were compared in pairs according to the pedigree analysis, and the coefficients of relationship, calculated using the ICIS program, were tested as independent (x) variables as a function of the dependent variables (y). In the case of genetic markers the dependent variables were calculated by comparing pairs of lines on the basis of the Dice similarity indexes. For the morphological descriptions, correlation coefficients calculated from the standardised values of the measured data using the SPSS statistical program, on the basis of square Euclidian distances, were compared with the independent variables.

RESULTS AND CONCLUSIONS

Pedigree analysis of maize inbred lines on the basis of line origin

On the dendrogram compiled on the basis of cluster analysis designed to model degrees of relationship, the classification of the lines into related groups corresponded well with the genetic background. The only exception was inbred line *Mv L20*, which was placed close to inbred line *Mv L17* on the dendrogram, despite

being of 50 % ISSS origin. However, since the other 50 % of its pedigree included a *W 117 × B 37* cross, also present in the pedigree of line *Mv L17*, its placement next to line *Mv L17* is genetically justified.

Polymorphism analysis

Polymorphism analysis on the basis of morphological descriptions

According to the results of analysis of variance, the environment had no significant effect on certain traits, such as the anthocyanine coloration of the spikelet husk, the shape of the ear, the kernel type, the colour of the kernel crown, the number of kernel rows, or the anthocyanine coloration of the cob. The anthocyanine coloration of the anther, the spikelet ring and the stigma, the degree of fertilisation of the ear and the width of the kernels exhibited only a slight dependence on the year (significant at the 10 % level of probability). The environment had a great influence on the kernel colour, where the data were significantly different at the 1 % level of probability. The remaining traits were significantly affected by the environment (p = 0.1 %).

On the basis of the UPOV TG 2/6 (1994) application form, two lines, Mo 17 Mv and its isogenic derivative, Mo 17 wx, exhibited such strong similarity that they could not be distinguished on the basis of the fundamental traits listed on the form.

The PIC values expressing the degree of polymorphism ranged from 0.48 to 0.82, with a mean value of 0.59. The highest PIC values were obtained for the anthocyanine coloration of the cob (0.82), tasselling (0.81) and silking (0.80).

Polymorphism analysis on the basis of isoenzyme patterns

Thirteen of the 15 enzyme loci examined exhibited polymorphism on the basis of isoenzyme patterns. Of the 35 alleles possible at these loci in all, 29 were found in the tested lines, equivalent to a mean value of 2.2 alleles/locus.

When determining polymorphism on the basis of isoenzyme patterns, the 46 lines were found to form 29 different gel electrophoresis groups, indicating that some of the lines could not be distinguished from each other. In the majority of cases, this similarity was due to the relationship between the lines and could be attributed to their having the same genetic background. In a few cases, however, the similarity between the isoenzyme patterns could not be explained on the basis of pedigree.

On the basis of the isoenzyme pattern 18 lines exhibited distinct gel electrophoresis patterns, manifested not so much in the appearance of unique alleles, as in distinct combinations of alleles at the polymorphic loci.

A similar picture was obtained in the case of Dice similarity indexes. The highest value was obtained for lines exhibiting identical enzyme patterns (Dice index = 1), which for most lines could be explained by the genetic background. The lowest value was obtained between *Mv L2* and *F7* (0.38), which corresponds well with the origin of the lines, which are not related to each other at all.

The two Lancaster lines, *Mo 17 Mv* and its isogenic variant, *Mo 17 wx*, could not be distinguished from each other on the basis of isoenzyme pattern either.

In order to determine the frequency of the isoenzyme alleles, the data of 294 inbred lines bred in Martonvásár and those of 119 other inbred lines obtained within the framework of international cooperation were compared with the results published by *Goodman and Stuber* (1983), who examined the allele frequency in 250 maize inbred lines.

The allele frequency of the Martonvásár lines was found to differ for some enzymes compared with both the foreign lines and the data in the literature. The null allele at the Mdh1 locus, for example, was only detected in the Martonvásár lines, while the null allele at the Idh1 locus was not detected in the foreign lines, but was reported in the literature to occur with a frequency of 0.23 %. A significant difference was detected in the frequency of allele 18 at the Mdh3 locus, allele 6 at the Idh1 locus, allele 6 at the Idh2 locus, allele 2.8 at the Pgd2 locus and allele 1 at the Pgm2 locus, all of which were found with lower frequency in the literature than in either the Martonvásár or the foreign lines. On the other hand, allele 15 at the Mdh5 locus,

allele 4 at the Idh2 locus and allele 2 at the Pgd1 locus were observed with significantly greater frequency according to *Goodman and Stuber* (1983).

The PIC (polymorphic index content) values expressing the degree of polymorphism ranged from 0.04 to 0.55 for the isoenzyme loci, with a mean value of 0.27. Among the polymorphic enzyme loci the lowest values were found for loci Mdh3, Mdh5 and Pgm1 (PIC = 0.04). The greatest amount of information with respect to the presence of polymorphism was provided by loci Pgm2 and Acp1, but even these only had PIC values of 0.55. The extremely low mean value and the fact that even the highest value was not very high indicate that the efficiency of isoenzyme analysis in demonstrating polymorphism is extremely limited.

Polymorphism analysis using PCR-based markers (RAPD, SSR)

Analysis involving RAPD and gene-linked microsatellite primer pairs was effective in detecting polymorphism in all the maize inbred lines tested.

Of the 20 RAPD primers tested, five did not give clear results even after several replications.

Three primers exhibited a 100 % monomorphic pattern. Other RAPD primers exhibited extremely selective polymorphism between the lines. Finally, 12 RAPD primers were chosen, with a total of 93 fragments, of which 54 (averaging 4.5 fragments/primer) gave a reliable polymorphic pattern in all the replications.

The Dice similarity index was high (0.8–0.9) in several related groups (e.g. ISSS, Lancaster, the early Canadian lines), which can be explained by the origin of the lines. The lowest Dice indexes calculated during the RAPD analysis were obtained between lines which, in the light of their genetic background, were known to be unrelated to each other. Some lines had low Dice indexes despite the fact that they were genetically related. The polymorphism between the two lines exhibiting the greatest similarity, *CM 105* and *CM 108* (Dice similarity coefficient: 0.91), was due to differences in two fragments on each of three primers.

In analysis based on gene-linked microsatellite primer pairs, three of the 20 primer pairs did not give PCR patterns suitable for evaluation, as the pattern was 100% monomorphic for two of them and almost 98% monomorphic for the third.

Although two of the primer pairs exhibited polymorphism, the patterns could not be identified because the differences between the fragments were extremely small, amounting to only a few base pairs. Several primer pairs gave extremely selective polymorphism between the lines. The data matrix was finally compiled on the basis of 71 polymorphic fragments from 11 primer pairs, representing 6.4 fragments/primer.

The selective polymorphism characteristic of the SSR markers was reflected in the Dice indexes, which had much lower values than in the case of isoenzyme patterns or RAPD analysis. High values were rarely found, and even these were lower than the highest values recorded for the other two methods. This indicates that the SSR markers provide an extremely selective analysis of polymorphism even with a low number of primers. The greatest similarity was found between lines *Mv L8* and *Mv L10* (Dice similarity coefficient: 0.85), the polymorphism of which was caused by two fragments from each of two microsatellite primer pairs.

An analysis of genetic markers enabled polymorphism to be detected for all the lines, which is particularly noteworthy in the case of isogenic lines (a line and its wx variant, or a line and its fertility restoring, rf, variant).

The two Lancaster lines, *Mo 17 Mv* and *Mo 17 wx*, which did not exhibit polymorphism on the basis of either morphological description or isoenzyme pattern, could be distinguished by means of both SSR and RAPD marker analysis.

The PIC (polymorphic index content) values expressing the degree of polymorphism were in line with the above observations.

The PIC values obtained for RAPD and SSR markers were the highest, ranging from 0.2–0.91 for RAPD, with an average of 0.61, and from 0.54–0.90 for SSR, with an average of 0.73. This indicates that the RAPD and SSR markers gave a reliable picture of polymorphism even with a relatively low number of primers.

Analysis of degrees of relationship

The analysis of degrees of relationship indicated that 46 was too large a number of inbred lines for genetic markers exhibiting little polymorphism, so the number of lines was reduced to 31. Even when the genetic markers were jointly evaluated, the dendrograms still only reflected the real genetic relationships in part, so the line collection was further reduced to include only two pairs of lines from each related group, chosen in such a way that they originated from each other or from a common ancestor. The new dendrograms depicted four groups on the basis of morphological description (Mindszentpusztai Yellow Dent, lines related to W 117, Iodent and ISSS lines originating from B 37) and four on the basis of isoenzyme pattern (ISSS lines originating from B 37, Lancaster, ISSS lines originating from B 14 and OP Lacaune lines). An analysis of the RAPD patterns showed six related groups (ISSS lines originating from B 14 and B 37, early Canadian, OP Lacaune, Co 125 and W 117 derivatives), which could be justified from the genetic background on the basis of the pedigree.

Even in this restricted collection of lines, the dendrograms compiled on the basis of the SSR markers exhibited an extremely heterogeneous picture. This is probably due to the fact that SSR markers bind to hypervariable regions of the genome, making them ideally suitable for the selective detection of polymorphism, but preventing the degree of relationship from being determined unless a large number of primers are analysed.

Cluster analysis based on the joint evaluation of the genetic markers placed all the lines in related groups justified by their genetic background.

The next step in the investigations was to determine what combination of morphological description and genetic markers would enable the lines to be grouped according to degrees of relationship, and the minimum number of markers required to make this classification reliable.

Among the laboratory analyses, the determination of enzyme patterns is the quickest, simplest and cheapest, so it is worth taking this as the basis. The measured

and scored morphological data routinely recorded in field experiments should also be utilised. These two experimental systems should then be supplemented with various genetic markers, increasing the number of primers step by step.

On the basis of morphological descriptions and enzyme patterns the related groups reflected the pedigree, with the exception of lines derived from Mindszentpusztai Yellow Dent.

The analysis was gradually expanded to include the results obtained with genetic markers until the groups reflected the real degrees of relationship. If all pairs of lines were to be separated correctly on the dendrograms, according to their pedigrees, it proved necessary to include a further three well-chosen RAPD primers exhibiting selective polymorphism.

The joint analysis of biochemical and genetic markers (isoenzyme, RAPD, SSR) placed all the lines in their correct related groups. This means that when a small number of lines are examined, genetic markers alone, without morphological descriptions, are sufficient to determine degrees of relationship.

Lines which cannot be classified in any of the major related groups were located on the dendrogram in positions corresponding to breeding experience, as they were grouped with lines exhibiting similarity in terms of either morphological traits, earliness or combining ability.

Linear regression analysis of the correlation between pedigree data, morphological descriptions, and biochemical and genetic markers

In the investigations the members of the three major related groups (Lancaster, ISSS, Iodent) were subjected to linear regression analysis.

The best correlation between the genetic distances obtained in pedigree analysis, reflecting the true genetic relationships, and those calculated using the SPSS program, based on a comparison of pairs of lines, was achieved through the joint evaluation of biochemical and genetic markers. In this case the correlation coefficient was $r = 0.81$, indicating a close correlation, compared with the medium values

obtained for morphological description (r = 0.48), isoenzyme patterns (r = 0.57), RAPD analysis (r = 0.64) and SSR analysis (r = 0,46) alone.

All the members of the three major related groups were also placed in groups reflecting their pedigrees when cluster analysis was carried out on the summarised data, indicating that the results of the two methods were in good agreement.

NEW SCIENTIFIC RESULTS

1. Polymorphism analysis based on morphological descriptions, isoenzyme patterns and DNA-based methods (RAPD and gene-linked microsatellite, or SSR, markers) was carried out on 46 maize inbred lines bred in Martonvásár or used as parents in the Martonvásár breeding programme. Such a comprehensive analysis of polymorphism has never before been carried out in Hungarian maize breeding.
2. Pedigree analysis was carried out on all 46 maize inbred lines, tracing the origin of each line back to the initial populations used in its development.
3. A comprehensive study was made on the frequency of the isoenzyme alleles occurring in the lines bred or utilised in the Martonvásár institute throughout its history. The allele frequency at the isoenzyme loci was determined for 294 maize inbred lines used as parents in the Martonvásár breeding programme and for 119 other lines obtained in the course of international cooperation, after which these data were compared with data from the literature on lines native to South America.
4. A system was developed whereby the data of morphological descriptions could be compared with the PIC (polymorphic index content) values used in polymorphism analysis based on genetic markers. It was then possible to determine which morphological traits were most suitable for detecting differences between maize varieties.
5. A limited line collection consisting of pairs of related lines with similar genetic backgrounds was subjected to cluster analysis, the results of which were used to set up a system making it possible to determine what combination of morphological descriptions and genetic markers was required to demonstrate genetic relationships reflecting the pedigree data. It was found that, if morphological descriptions and isoenzyme patterns, which are routinely established data, are taken as the basis, it is sufficient to include an analysis of

three well-chosen RAPD primers exhibiting selective polymorphism to achieve results reflecting the true genetic background known from the pedigree.

6. When the number of related groups was restricted and the data were jointly evaluated, all the members of the three main related groups (Reid Yellow Dent, Iodent, Lancaster) were placed in the group reflecting their pedigrees.
7. If morphological descriptions are not available, a joint analysis of genetic and biochemical (isoenzyme) markers proved to be adequate to give a precise indication of genetic relationships.
8. These conclusions were confirmed by the results of linear regression analysis, indicating that the relationship coefficients determined by pedigree analysis are more closely correlated with those estimated by the joint processing of genetic markers than with those estimated using the various methods separately.
9. A study of where lines not classified in any of the major related groups were located on the dendrogram compiled for the restricted line collection indicated that all five lines were placed in groups justified by breeding experience, being adjacent to lines similar to them in terms of either morphological traits or earliness or their combining ability with other lines.

PUBLICATIONS ON THE SUBJECT OF THE THESIS

Papers in scientific journals

E. Nagy, G. Gyulai, Z. Szabó, Z. Hegyi, L. C. Marton (2003): Application of morphological descriptions and genetic markers to analyse polymorphism and genetic relationships in maize (*Zea mays* L.) Acta Agronomica Hungarica 51, 257-265.

S. Záborszki, **E. Nagy**, C. Szőke (2002): Effect of seed treatment on the emergence of inbred lines of maize (*Zea mays* L.) Acta Agronomica Hungarica, 50, 359-369.

Záborszky S., **Nagy E.**, Berzy T., Pintér J. (2002): A vetőmagcsávázás és a vetésidő hatása a kukoricahibridek (Zea mays L.) kelésére és kezdeti fejlődésére. Növényvédelem, 38, 629-635.

Z. Gyenes-Hegyi, I. Pók, L. Kizmus, Z. Zsubori, **E. Nagy**, L.C. Marton (2002): Plant height and height of the main ear in maize (*Zea mays* L.) at different locations and different densities. Acta Agronomica Hungarica, 50, 75-84.

M. Pál, **E. Nagy** (2002): Chilling tolerance of maize. Acta Agronomica Hungarica, 50, 95-106.

Marton L. C., Kizmus L., **Nagy E.** (2000): A fuzáriumos magfertőzés hatása a kukorica (*Zea mays* L.) keléskori hidegtűrésére. Növénytermelés, 3, 261-272.

Nagy E. Gyulai G., Marton L. C. (2000): Kukorica beltenyésztett törzsek jellemzése genetikai markerekkel. Növénytermelés, 49, 587-597.

Marton L. C., Szundy T., **Nagy E.** (1997): A kukorica (*Zea mays* L.) fiatalkori hidegtűrésének értékelése hőmérsékleti gradiens kamrában. Növénytermelés, 46, 549-557.

Conference Proceedings

Nagy E., Gyulai G., Marton L. C. (1999): Genetikai markerek felhasználása a kukoricanemesítésben. Ötven éves a Magyar Tudományos Akadémia Mezőgazdasági Kutatóintézete. Kiadó: A Magyar Tudományos Akadémia Kutatóintézete, 131-135.

E. Nagy, G. Gyulai and L. C. Marton (1999): Use of genetic markers in maize breeding. 50th Anniversary of the Agricultural Research Institute of the Hungarian Academy of Sciences. Agricultural Research Institute, Martonvásár, pp. 135-139.

Marton L. Cs., Hadi G., Pintér J., **Nagy E.** (2003): A tőszám hatása a Mindszentpusztai Sárga Lófogú fajta szárszilárdságára. 50 éves a magyar hibridkukorica. Kiadó: A Magyar Tudományos Akadémia Kutatóintézete, 209-216.

Nagy E., Hegyi Zs., Marton L. Cs. (2003): Genetikai markerek a kukorica polimorfizmus és rokonsági vizsgálatokban. 50 éves a magyar hibridkukorica. Kiadó: A Magyar Tudományos Akadémia Kutatóintézete, 223-228.

Abstracts

Gyenesné Hegyi Z., Kizmus, L., **Nagy, E**., Marton, L. C. (2001): A kukorica címerágak számának és az egyedi produkciónak a vizsgálata eltérő ökológiai körülmények között. II. Növénytermesztési Tudományos Napok, Budapest

Nagy E., Gyulai G., Gyenesné Hegyi Z., Szabó Z., Marton L. C. (2002): Genetikai távolság becslése génkapcsolt mikroszatellita markerekkel kukoricában. VIII. Növénynemesítési Tudományos Napok, Budapest

Nagy E., Pál M. (2002): Alkalmazható módszerek a kukorica hidegtűrése kutatásában. VIII. Növénynemesítési Tudományos Napok, Budapest

Gyenesné Hegyi, Z., Pók, I., Zsubori, Z., Kizmus, L., **Nagy, E**., Marton, L. C. (2002): A kukorica (*Zea mays* L.) növénymagasságának és főcső eredési magasságának alakulása eltérő termőhelyeken. VIII. Növénynemesítési Tudományos Napok, Budapest

Oral Presentations

Nagy E., Marton L. C., Szundy T. (1998): Genetikai markerek a kukoricanemesítésben: martonvásári kukoricafajták genetikai távolságának meghatározása morfológiai tulajdonságok és izoenzim mintázat alapján. IV. Növénynemesítési Tudományos Napok, Budapest

Nagy E., Gyulai G., Marton L. C. (1999): Genetikai markerek felhasználása a kukoricanemesítésben. Ötven éves a Magyar Tudományos Akadémia Mezőgazdasági Kutatóintézete