Original Paper Open Access

# Assesment of genetic diversity and relationships among maize inbred lines developed in Italy

**Alessia Losa, Hans Hartings*, Alberto Verderio, Mario Motto**

Consiglio per la Ricerca e la Sperimentazione in Agricoltura, Unità di Ricerca per la Maiscoltura, via Stezzano 24, 24126 Bergamo, Italy
*Corresponding author: E-mail: hans.hartings@entecra.it

## Abstract

The genetic diversity pattern of a sample of 144 maize inbred lines comprising 106 Italian entries, considered representative of the breeding material developed at the Bergamo Maize Breeding Station, and a sample of 38, mainly US Corn Belt based, reference lines was accessed using AFLP markers. A total of 811 polymorphic fragments were identified. Exploration of the variation disclosed by the lines by means of principal component analysis (PCA) and hierarchical clustering allowed their division into major heterotic groups. The obtained grouping of the inbred lines reflected pedigree information and resulted in the identification of major clusters derived from Lancaster Sure Crop (LSC), Iowa Stiff Stock Synthetic (BSSS), and miscellaneous heterotic breeding material. AMOVA statistics, performed on the established genetic structure, revealed a high proportion of variance between individuals and among populations stressing the high polymorphic nature of the maize pool analyzed. Regarding population structures, the genetic distance among populations ($F_{ST}$ = 0.50 ± 0.1) and the degree of inbreeding within groups ($F_{SC}$ = 0.46 ± 0.1) did not diverge significantly, while both significantly differed from the degree of relatedness between markers within groups ($F_{CT}$ = 0.06 ± 0.04). In conclusion, the results presented indicate that AFLPs are useful in assigning inbred lines to heterotic groups and for superior line development with the aim to maximize heterosis and consequently yield performance.

**Keywords**: breeding groups, inbred lines, molecular analysis, *Zea mays* L.

## Introduction

Maize (*Zea mays* L.) is an important food and animal feed worldwide, and occupies a relevant place in the world economy and trade as an industrial grain crop (White and Johnson, 2003). Although maize is produced primarily (80%) as an energy crop for animal feeding, specialized versions for human consumption and industrial use are available. Moreover, it is a model system for the study of genetics, evolution, and domestication.

Detailed knowledge of the relationships between maize breeding lines is important not only for parental selection but also for genetic analysis and breeding system design (Hallauer et al, 1988). In fact, this information is useful in planning crosses for hybrid and line development, in assigning lines to heterotic groups, as well as in plant variety protection. Moreover, effective use of maize germplasm in breeding programs requires accurate characterization of line performance and line relationships to other germplasm. When developing breeding populations, maize breeders should choose parents that i) exhibit superior performance for the traits of interest, ii) maximize within-population variance for the traits of interest, and iii) preserve heterotic patterns for maximum heterosis in hybrid development. To this scope, breeders require phenotypic data on potential parents and an understanding of the relationships among these lines. Additionally, knowledge of the genetic relationships among breeding materials may help to prevent the risk of increasing uniformity in the elite germplasms and may also ensure long-term selection gains (reviewed in Pollack, 2003). There is evidence indicating that genetic diversity within maize is decreasing at an alarming rate because of modern hybrids and agricultural practices (Duvick, 2004; Reif, 2005). In this respect, maize breeders have become more aware of the necessity to preserve genetic diversity and associated phenotypic variability.

Diversity analysis of germplasm collections can be obtained from pedigree and test cross data at morphological, geographical, molecular (DNA, sequence, gene), and functional levels (e.g. Buckler et al, 2006; Messing and Dooner, 2006; Springer et al, 2009; Gilliland et al, 2000). In particular, molecular markers have been widely used in maize genetic diversity studies for the: i) analysis of genotype frequencies for identification of deviations at individual loci and for characterization of molecular variation within or between populations, ii) construction of “phylogenetic” trees and determination of heterotic groups, and iii) analysis of correlation between genetic distance and hybrid performance, heterosis (when the hybrid shows


RECEIVED 29/05/2011

vigor superior to its parents), and specific combining ability (reviewed in Xu et al, 2009).

Among molecular markers, amplified fragment length polymorphisms (AFLPs; Vos et al. 1995) appear very useful for the analysis of within-species variation since they allow the rapid acquisition of a large amount of genetic information, due to the capability to simultaneously identify a large number of amplification products (reviewed in Bonin et al, 2007). The AFLP technique has been largely used in maize to construct genetic maps, to study phylogenic relationships, and measure genetic diversity (e.g., Lubberstedt et al, 2000; Ajmone-Marsan et al, 2001; Stich et al, 2006, and references therein).

Accordingly, the aims of this study were to i) monitor the genetic variation, as sampled by AFLP, in a collection of inbred lines developed in Italy in the last decades by the Bergamo maize Station; ii) determine the level of genetic diversity found within and between different heterotic groups; iii) explore the usefulness of AFLPs for assigning inbred lines to heterotic groups.

## Materials and Methods

### *Plant material*

One hundred and forty four maize accessions, chosen to represent diverse germplasms selected in climatically temperate locations, were used as the experimental material. The majority of these inbreds have been used in previous decades for the production of hybrid seed in Italian breeding programs. Among these inbreds, 106 were developed at the Maize Breeding Station at Bergamo (Italy), while the remaining, a group of 38 historically highly selected, and elite inbred lines, represented a broad range of diversity from the U.S. Corn Belt and Argentina (A69Y), and were included for comparison. The inbred lines considered together with their pedigree information are listed in Table 1, while the reference lines included in this study are summarized in Table 2. Of the 144 inbred lines analyzed, 69 (47.9%) belonged to the Iowa Stiff Stalk Synthetic (BSSS) heterotic group, 47 (32.7%) belonged to the Lancaster Sure Crop (LSC) group, and 28 (19.4%) were of independent origin. In this study, the LSC heterotic group will be used in a broad context including lines that either contain primary LSC germplasm or have a good combining ability towards lines within the BSSS heterotic group. All entries were grown in field trials at Bergamo in 2008-2010, using a randomized complete block design with three replications. Experimental plots consisted of four rows, each containing 25 plants, at a density of 57,000 plants/ha. Recommended crop-management techniques were applied.

### *Molecular analysis*

Thirty individuals were sampled from each inbred according to indications reported by Crossa et al (1993). Genomic DNA was extracted from shoots of 2-week-old germinated seedlings of each accession as described in Chittò et al (2000). Molecular genotyping was carried out using the AFLP protocol according to Vos et al (1995), using either EcoRI or PstI as the rare cutting and MseI as the frequently cutting restriction enzyme. Briefly, upon DNA digestion, specific adaptors were ligated onto the digested DNA. Then, pre-amplification was performed with a primer carrying an adenosine as the selective nucleotide. Subsequently, amplification was achieved using 3 selective nucleotides for the EcoRI and MseI primers and two selective nucleotides for the PstI primers. Fourteen combinations of selective primers with a high polymorphic detection rate in maize were employed (Chittò et al, 2000) and are listed in Table 3. Autoradiographs were manually scored for major polymorphic bands, ignoring low signal fragments. Finally, a two-dimensional matrix was constructed, representing the absence/presence of each polymorphic fragment within the accessions considered. The nucleotide sequences of the AFLP primers used are available on demand.

### *Statistical analysis*

Manual scoring of autoradiographs, considering the presence (1) or absence (0) of bands in each combination of genotypes, allowed the construction of a binary matrix, which was used to determine the polymorphism information content (PIC=1- $\sum f_i^2$) of the AFLP markers. This value defines the probability that two alleles taken at random from a population can be distinguished using the marker in question (Smith et al, 1997). The AFLP technology produces dominant markers and only two states can be distinguished for each band. Hence, a maximum PIC value of 0.5 can be obtained.

The binary AFLP data were, furthermore, used to derive genetic similarities (GS) according to Nei and Li (1979): $GS_{ij} = 2N_{ij}/(N_i + N_j)$, where $N_{ij}$ is the total number of bands common to genotypes i and j, and $N_i$ and $N_j$ are the total number of bands present in genotypes i and j, respectively. Genetic distances were obtained from similarity values (GD = 1 – GS) and used to perform Neighbor-joining cluster analysis. A consensus tree was obtained by performing a bootstrap analysis (10,000 iterations) on the resulting dendrogram. All similarity and bootstrapping analyses were performed with the NTSYS-PC (Rohlf, 1993) and PAST (Hammer, 2001) software packages. Furthermore, principal component analyses (PCA) were performed on the similarity matrix using the STATISTICA software suite (StatSoft).

Analyses of molecular variance (AMOVA) were performed on the data set to partition the observed variation across the accessions considered using Arlequin version 3.5 (Excoffier et al, 2010). This software was furthermore employed to compute the degree of inbreeding within groups ($F_{SC}$), the degree of relatedness between markers within groups ($F_{CT}$), and the unbiased estimates of Wright's fixation index ($F_{ST}$) according to Weir and Cockerham (1984).

Table 1 - Summary of inbred lines analyzed and respective pedigrees.

| Inbred | Pedigree | BG[1] | Inbred | Pedigree | BG |
|---|---|---|---|---|---|
| Lo3 | Nostrano isola | M | Lo1131 | Lo9042 x Lo951 | B |
| Lo863 | Nostrano isola | M | Lo1137 | P3343 | B |
| Lo876 | Lo876o2 x BSSS | B | Lo1140 | Synthetic WF9 | L |
| Lo881 | Synthetic C103 | L | Lo1141 | Synthetic B37 | B |
| Lo902 | Mo172 x P3780A | L | Lo1142 | Lo983 x Lo1063 | L |
| Lo903 | B732 x B37 | B | Lo1154 | Lo924 x Lo1063 | M |
| Lo904 | B732 x B37 | B | Lo1156 | P3245 | B |
| Lo924 | H992 x Mo17 | L | Lo1157 | P3245 | B |
| Lo932 | Synthetic BS5 | M | Lo1158 | P3245 | B |
| Lo933 | Synthetic GD x BS5 | M | Lo1159 | P3245 | B |
| Lo937 | Synthetic BS5 | M | Lo1160 | Lo1061 x Lo1090 | L |
| Lo944 | Synthetic BS5 | M | Lo1162 | Lo1061 x Lo1090 | L |
| Lo950 | P3183 | B | Lo1166 | Lo9242 x Lo1063 | L |
| Lo951 | P3183 | B | Lo1167 | Lo904 x LA47677 | B |
| Lo960 | P3183 | B | Lo1168 | Lo1063 x P3374 | B |
| Lo964 | P3183 | B | Lo1169 | Lo904 x Lo1067 | B |
| Lo976 | Mo172 x LA4317 | L | Lo1170 | Lo10412 x Lo1063 | B |
| Lo986 | Synthetic Ostrinia | M | Lo1171 | LA47678 x P3979 | B |
| Lo999 | B73 x Teosinte | B | Lo1172 | Lo1059 x Lo863 | L |
| Lo1010 | B372 x VA885 | M | Lo1173 | Lo1094 | B |
| Lo1016 | P3369A x Lo876o2 | B | Lo1176 | Tosca | B |
| Lo1025 | B732 x Lo876 | B | Lo1180 | Lo1074 x P3539 | L |
| Lo1026 | B732 x Lo876o2 | B | Lo1182 | Lo1059 x Lo1077 | L |
| Lo1035 | P3183 x Va59 | L | Lo1187A | Lo904 x Lo1095 | B |
| Lo1038 | P3183 x Va93 | B | Lo1187D | Lo904 x Lo1095 | B |
| Lo1053 | Lo950 x Lo951 | B | Lo1189 | LA47678 x P3245 | B |
| Lo1054 | Lo950 x Lo951 | B | Lo1199 | Lo1086 x Lo1094 | B |
| Lo1055 | Lo950 x Lo951 | B | Lo1203 | Lo1095 x Lo1125 | B |
| Lo1056 | Lo8812 x Lo964 | L | Lo1223 | Lo904 x LA59282 | B |
| Lo1059 | P3297 | L | Lo1241 | Lo1067 x Lo1125 | B |
| Lo1061 | P3297 | L | Lo1242 | Lo1124 x Lo1096 | L |
| Lo1063 | P3297 | L | Lo1246 | Lo1142 | L |
| Lo1064 | Lo876o2 x N7A | B | Lo1251 | Lo1094 x Lo1159 | B |
| Lo1066 | Lo876o2 x A641 | B | Lo1253 | Lo1094 x Lo1159 | B |
| Lo1067 | P3780A x Lo876o2 | B | Lo1255 | Lo1095 x Lo1125 | B |
| Lo1074 | Synthetic MP | B | Lo1260 | P3394 | L |
| Lo1076 | P3297 | L | Lo1261 | Lo904 x Lo1087 | B |
| Lo1077 | P3540 | L | Lo1263 | Lo904 x Lo1125 | B |
| Lo1086 | Lo9042 x Lo951 | B | Lo1265 | Lo904 x Lo1125 | B |
| Lo1087 | Lo9512 x Lo904 | B | Lo1266 | P3394 x Lo1077 | L |
| Lo1090 | Lo8812 x Lo964 | L | Lo1270 | Lo1056 x Latina | L |
| Lo1094 | Synthetic BGSF | B | Lo1273 | Lo1095 x Lo1123 | B |
| Lo1095 | P3189 | B | Lo1274 | Lo1124 x Cecilia | L |
| Lo1096 | P3540 | L | Lo1279 | Lo1101 x Lo1125 | B |
| Lo1101 | Lo904 x I26847 | B | Lo1280 | Lo1061 x Lo1124 | L |
| Lo1106 | Synt. SSS Elite Bg | B | Lo1282 | P3374 | L |
| Lo1123 | A632 x P3540 | B | Lo1284 | P3374 | L |
| Lo1124 | Lo924 x Lo1063 | L | Lo1288 | Lo1130 x Lo1124 | L |
| Lo1125 | P3374 | B | Lo1290 | Lo1124 x Lo1158 | L |
| Lo1126 | Lo993 x Lo1063 | L | Lo1292 | Lo1061 x Lo1124 | L |
| Lo1127 | CD1 x P3551 | B | Lo1297 | Lo1131 x Lo1123 | B |
| Lo1128 | P3374 | B | Lo1301 | Lo1173 x Lo1101 | B |
| Lo1129 | P3394 | L | Lo1322 | Lo1208 x Lo1059 | L |

[1]Background - B: BSSS = Iowa Stiff Stalk Synthetic; L: LSC = Lancaster Sure Crop; M = miscellaneous

Table 2 - Summary of reference lines analyzed and respective pedigrees.

| Inbred | Pedigree | BG[1] |
|---|---|---|
| A619 | (A171 x Oh43) x Oh43 | L |
| A632 | (Mt42 x B14) x B143 | B |
| A69Y | Plata argentina | M |
| A71 | Funk Yellow Dent | M |
| B14 | Iowa Stiff Stalk Synthetic | B |
| B37 | Iowa Stiff Stalk Synthetic | B |
| B57 | Midland | M |
| B73 | Iowa Stiff Stalk Synthetic | B |
| B84 | BSSS13 | B |
| B89 | BSSS(R)C7-84 | B |
| B103 | NT Pool 41-C15-9-1 | B |
| C103 | Lancaster Sure Crop | L |
| CI187-2 | CI187 x B2 rec.blight rest | L |
| FR5 | Oh07 Sister | M |
| H55 | Hy2 x Mo21A | M |
| H95 | Oh43 x CI90A | L |
| H96 | H55 x H56 | M |
| H99 | Illinois Synthetic 60C | L |
| K55w | Pride of Saline | M |
| Mo17 | CI187-2 x C103 | L |
| N6 | Hayes Golden | M |
| N22A | N22 Outcross | M |
| N28 | SSS1 Synthetic | B |
| NC250 | B372 x Nigeria Comp ARb | B |
| NC260 | Mo443 x Mo17 | L |
| Oh07 | CI540 x IIIL | M |
| Oh33 | Clarage | M |
| Oh40B | Lancaster Composite | L |
| OH43 | Oh40B x W8 | M |
| OS420 | Osterland Yellow Dent | M |
| Pa91 | (WF9 x Oh40B) x (38-112 x L317) | M |
| T8 | Jarvis Golden Prolific | M |
| Va26 | Oh43 x K155 | L |
| Va59 | C1032 x (T8 x K4) | L |
| Va85 | Virginia Long Ear Synthetic | M |
| W153 | Ia1532 x W8 | M |
| W64A | Wf9 x CI187-2 | M |
| WF9 | Wilson Farm Reid | M |

[1]Background - B: BSSS = Iowa Stiff Stalk Synthetic; L: LSC = Lancaster Sure Crop; M = miscellaneous

## Results

AFLP analyses of the inbred lines produced stable and repeatable profiles, which allowed us to unequivocally fingerprint each inbred tested. For each accession, approximately 150-200 amplified fragments could be visualized in each AFLP run depending on the primer pair employed. In total, the 14 primer pairs used (10 E/M and 4 P/M; Table 3) produced 811 polymorphic AFLP bands on the 144 inbreds analyzed. Although only major polymorphisms were scored as described, an average of over 57 markers could be obtained for each primer combination, confirming the power of AFLP analysis in DNA profiling of maize, with substantial polymorphisms between varieties. The number of markers per primer pair ranged from 35 (primer combination E33M47) to 65 (primer combination E38M51).

The existence of 811 AFLP loci appeared sufficient to investigate the genetic structure of the 144 populations, i.e. relatively distantly related entities. The polymorphism information content (PIC) measured 0.34 ± 0.14 on average, while individual values ranged from 0.02 to 0.50. Approximately 44% of the loci used (355 out of 811) had a PIC value exceeding 0.3, demonstrating the good discriminatory power of the markers identified (Figure 1) and suggesting that considerable variation between inbreds is detectable with AFLP markers.

The scored AFLP profiles were used to calculate a matrix of genetic similarities (GS) according to Nei and Li (1979). GS distances were subsequently transformed in genetic distance (GD) values. GDs ranged from 0.115 for inbred lines Lo1094 and Lo1173, both belonging to the BSSS heterotic group, to 0.613 for inbreds Lo976 and Lo1169, derived from Mo17 and B73, respectively. An average GD of 0.278 ± 0.084 was calculated for the entire data set. The minimum and maximum GD values observed are in good agreement with previous data regarding a subset of the accessions analyzed in this work (Chittò et al, 2000).

In order to investigate the distribution of variability across the group of inbred lines considered, Principle Components Analysis (PCA) was performed on the calculated GD values. Figure 2 represents a graphical distribution of the landraces in a plain defined by the first two PCs, which accounted for 27.8% and 17.9% of the total variability, respectively. This combination of components reveals a clear distribution of the accessions across the plain considered and evinces a good separation of the classical breeding groups present within the maize lines considered. In particu-

Table 3 - AFLP primer combinations used in this study. Primer codes and 3' selective nucleotides are given.

| EcoRI | 3' | MseI | 3' |
|---|---|---|---|
| E32 | AAC | M50 | CAT |
| E32 | AAC | M60 | CTC |
| E33 | AAG | M47 | CAA |
| E33 | AAG | M50 | CAT |
| E33 | AAG | M61 | CTG |
| E35 | ACA | M49 | CAG |
| E35 | ACA | M50 | CAT |
| E35 | ACA | M58 | CGT |
| E38 | ACT | M47 | CAA |
| E38 | ACT | M51 | CCA |
| **PstI** | **3'** | **MseI** | **3'** |
| P12 | AC | M49 | CAG |
| P12 | AC | M50 | CAT |
| P13 | AG | M50 | CAT |
| P13 | AG | M61 | CTG |

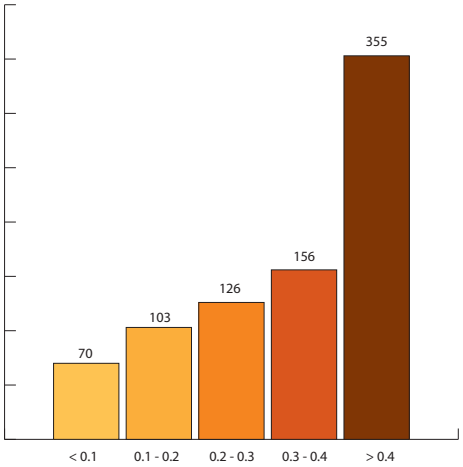


Figure 1 - Distribution of PIC values. Markers were divided into groups based on PIC values. The number of markers contained in each group is indicated.

lar, the first PC determines a horizontal spread of the BSSS, LSC, and miscellaneous heterotic breeding materials with a substantial separation of the latter two. The second PC distinguishes the BSSS breeding material, collocated mainly below the PC axis from the remaining heterotic material, represented above the axis. Hence, the variability present within the dataset produced by the AFLP primer-enzyme combinations was important in defining the major heterotic groups, separating lines with a BSSS, LSC, or miscellaneous origin.

Cluster analysis was used to further investigate the inter-relationships among the 144 inbred lines. Dendrograms are an effective mean of quantifying patterns in genetic distances between close neighbors (Mumm and Dudley, 1994). Therefore, the obtained GS matrix was used to perform a hierarchical clustering analysis by means of the neighbor-joining method. The resulting phylogenetic tree was subjected to bootstrap analyses using 10,000 iterations. Figure 3 shows the result of these analyses, distinguishing those branches exhibiting bootstrap consensus values greater than 67%. This graphical representation reveals three major clusters, i.e. a BSSS (group I), an LSC (group II) and a miscellaneous cluster (group III) of inbred lines exhibiting a large degree of differentiation within each cluster. Thus, the predominant heterotic groups and important subgroups within each heterotic group were represented in the dendrogram. This was, similarly, suggested by the presence of at least 10 predominant sub-clusters, represented by the following lines: B73, B37, Lo1077, Mo17, H95, A632, Oh33, A619, C103, and Oh43, each clearly separated in the phylogenetic tree, confirming the highly polymorphic nature of the inbred pool considered.

Within the BSSS cluster, two distinctive sub-clusters were identified. The first sub-cluster is formed around the Lo950 and Lo951 inbred lines, which were selfed out from P3183, a commercial hybrid. Members of this group showed reasonable genetic similarity and are associated with lines with B73 and B37 backgrounds such as Lo904. The second sub-cluster was mainly formed by inbreds derived from commercial hybrids as second cycle improvements after crossing with previous elite Lo inbred line germplasm (e.g. Lo1058-, Lo1180-, and Lo1128-types).

Similarly, on the LSC side two prevalent clusters appeared. Within the first sub-cluster were historical inbred lines derived from synthetic Corn Belt varieties, along with B84- and H55-derived distinctive groups. In the second sub-cluster Va59, T8, and C103 were highly clustered and distantly merged with Lo881 types, with inbreds selected from a C103 synthetic, a narrow based gene pool derived from intercrossing C103 derived lines (Bertolini et al, 1991), and with second cycle improved inbred lines derived from Lo881. In this second sub-cluster, the Oh43 inbred was highly clustered with Lo1126, an inbred that was selfed out from a commercial hybrid (P3297). These inbreds further merged with CI187-2, and with two Lo lines (Lo902 and Lo976), originated from a synthetic Lancaster. Lo1035 and Lo1038, derived from the elite lines Va59 and Va53, and a commercial hybrid (P3183) were highly clustered and merged with C103 types.

Within the third major cluster, at least 7 related sub-clusters were identified. This group included reference inbreds such as Pa91, LSC, N6A, Mo17, H95, NC250, Oh40B, and N28, derived from various synthetics or populations of the U.S. Corn Belt, and groups of Lo inbred lines (Lo1242, Lo1322, Lo1297, Lo1265), selected from commercial hybrids or as second cycle improvements of Lo inbred lines and commercial hybrids. Furthermore, an independent sub-group formed by Nostrano dell'Isola-types (Lo3 and Lo881), BS5, a synthetic variety from Iowa re-selected in Italy, and derived types (Lo932, Lo934, and Lo937) was apparent. All of these merged with the Lo1154 and Lo1189 lines, containing commercial hybrid germplasm, and with Va56, B14, and Oh33. In Table 4, the inbreds representing the major groups identified in this study, as well as their disclosed subgroups are summarized.

For data with a hierarchical structure, analysis of molecular variance (AMOVA) allows the study of patterns of genetic variation within and between groups through the examination of variance. This assay can be extended to evaluate molecular marker data even in the absence of replicated values for samples (Law et al, 1998). An AMOVA analysis based on genetic distances derived from the obtained AFLP data as visualized through the clustering of the Italian and reference inbred lines considered in this study (Tables 1 and 2), was performed. Clusters were used to recompose, in broad terms, BSSS, LSC, and miscellaneous heterotic groups. Amalgamation was performed us-

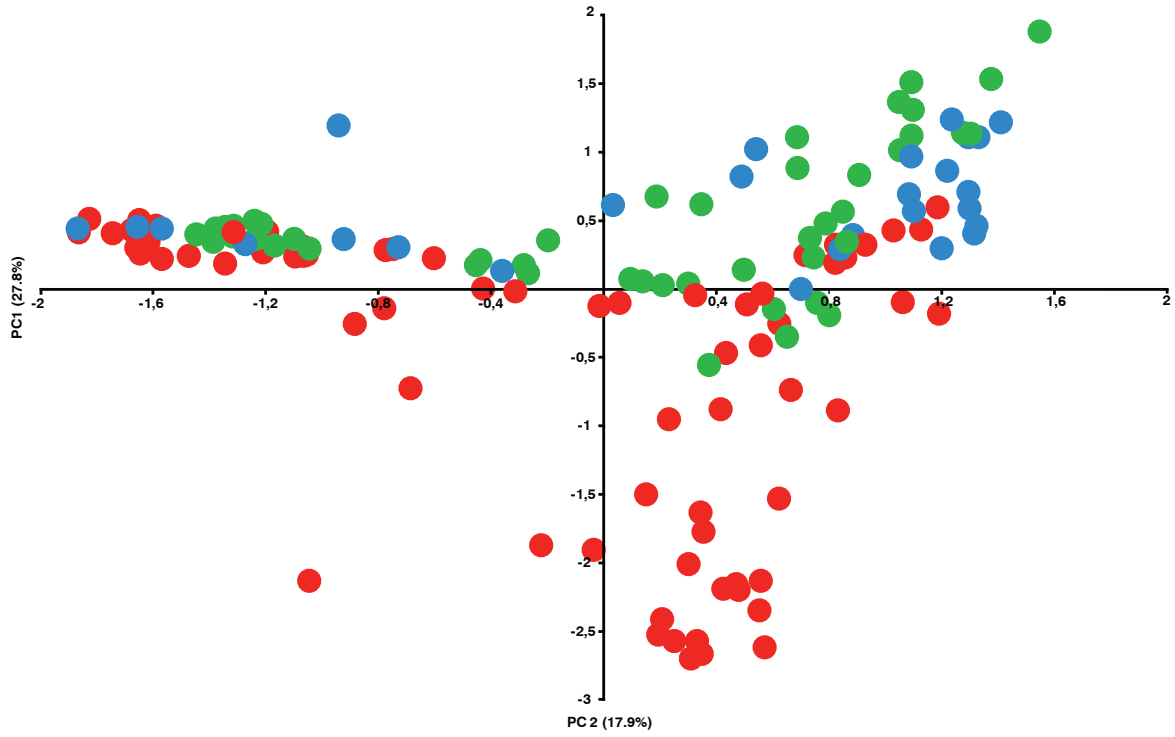


Figure 2 - Principal component analysis of 144 Italian and reference maize inbred lines based on AFLP markers. Accessions are color coded following their assumed pedigree as reported in Tables 1 and 2 (red = BSSS; green = LSC; blue = independent).

ing both a small number of larger clusters and a larger number of clusters of reduced size. In both cases, the within-population (clusters) components of variance dominated the AMOVA, accounting for approximately 50% of the variation. Conversely, a low level (6%) of differentiation was detected among groups (Table 5). Changes in the grouping pattern applied had no significant effect on the distribution of variation. Furthermore, the genetic distance between populations ($F_{ST}$ = 0.50 ± 0.1) did not significantly differ from the extent of inbreeding within groups ($F_{SC}$ = 0.46 ± 0.1). The degree of relatedness between markers within groups ($F_{CT}$ = 0.06 ± 0.04) was significantly low (Table 5).

## Discussion

In breeding programs, information on genetic relationships within and between species is used for organizing germplasm collections, identifying heterotic groups within crops and selecting parents for purposes of crossing. In this study we present a fingerprinting analysis, based on molecular markers, allowing the identification of genetic variation and the relationships among accessions at the molecular level. This method allows the simultaneous detection of numerous variable regions with a single probe, yielding an individual specific banding pattern in different organisms. DNA fingerprinting has been used for a variety of purposes, such as parentage testing, individual identification, and the acceleration of breeding programs (Xu et al, 2009).

In the current study, DNA fingerprinting was used to analyze the genetic diversity patterns in a sample of 106 Italian inbred lines, considered representative of the breeding material developed at the Bergamo Maize Station together with a set of historical elite lines encompassing the major maize heterotic groups. In particular, the reference lines supplied a basis of genetic diversity to which the Italian inbred lines were related in the evaluation of their relative genetic relationships. The Italian inbred lines presented here have been released to the maize breeding community over a time span of 25 year from 1981 to 2006, while the reference lines are distinctive for the major heterotic groups available, supplying the basis of genetic diversity to which the Italian inbred lines were correlated in the evaluation of their relative genetic relationships.

The use of AFLP markers in diversity analysis has been frequently criticized as they represent markers of unknown genomic distribution, which could, hence, genetically cluster without providing a genome-wide coverage. It has been argued that ample coverage represents an essential asset for the examination of genetic diversity (Karp et al, 1997). To partially avoid this limitation, a subset of the AFLP results was obtained with the use of the methylation-sensitive PstI restriction enzyme. The use of methylation-sensitive

Table 4 - The major heterotic groups identified in this study.

| BSSS1 | | BSSS2 | LSC1 | | LSC2 | | Misc1 | | Misc2 |
|---|---|---|---|---|---|---|---|---|---|
| **B73**[a] | **B37** | **Lo1063** | **C103** | **Oh43** | **A619** | **Wf9** | **Pa91** | **Oh40B** | **B14** |
| Lo903 | Lo876 | Lo1059 | T8 | CI187-2 | A69Y | A71 | Mo17 | A632 | Oh33 |
| Lo904 | Lo999 | Lo1061 | Va59 | OS420 | B84 | W153 | N22A | B103 | Va26 |
| Lo950 | Lo1016 | Lo1074 | Lo881 | Va85 | FR5 | W64A | Lo1025 | B57 | Lo3 |
| Lo951 | Lo1055 | Lo1076 | Lo1035 | Lo902 | H55 | | Lo1026 | B89 | Lo863 |
| Lo960 | Lo1064 | Lo1077 | Lo1038 | Lo924 | H99 | | Lo1131 | H95 | Lo932 |
| Lo964 | Lo1066 | Lo1095 | Lo1056 | Lo976 | N6 | | Lo1187A | H96 | Lo937 |
| Lo1053 | Lo1067 | Lo1096 | Lo1090 | Lo1124 | NC260 | | Lo1187D | K55w | Lo944 |
| Lo1054 | Lo1141 | Lo1123 | Lo1140 | Lo1126 | Oh07 | | Lo1223 | N28 | Lo1154 |
| Lo1086 | Lo1169 | Lo1125 | | Lo1142 | Lo933 | | Lo1261 | NC250 | Lo1189 |
| Lo1087 | | Lo1128 | | Lo1166 | Lo986 | | Lo1263 | Lo1129 | |
| Lo1094 | | Lo1156 | | Lo1170 | Lo1010 | | Lo1279 | Lo1171 | |
| Lo1101 | | Lo1157 | | Lo1292 | Lo1255 | | Lo1301 | Lo1199 | |
| Lo1106 | | Lo1158 | | | Lo1260 | | | Lo1203 | |
| Lo1127 | | Lo1159 | | | Lo1273 | | | Lo1241 | |
| Lo1137 | | Lo1160 | | | Lo1280 | | | Lo1242 | |
| Lo1167 | | Lo1162 | | | | | | Lo1246 | |
| Lo1173 | | Lo1168 | | | | | | Lo1251 | |
| | | Lo1172 | | | | | | Lo1253 | |
| | | Lo1176 | | | | | | Lo1265 | |
| | | Lo1180 | | | | | | Lo1266 | |
| | | Lo1182 | | | | | | Lo1270 | |
| | | | | | | | | Lo1274 | |
| | | | | | | | | Lo1282 | |
| | | | | | | | | Lo1284 | |
| | | | | | | | | Lo1288 | |
| | | | | | | | | Lo1290 | |
| | | | | | | | | Lo1297 | |
| | | | | | | | | Lo1322 | |
| 0.30[b] | 0.37 | 0.39 | 0.27 | 0.40 | 0.42 | 0.28 | 0.29 | 0.31 | 0.37 |
| 0.13[c] | 0.12 | 0.15 | 0.22 | 0.13 | 0.30 | 0.19 | 0.17 | 0.12 | 0.29 |
| 0.52[d] | 0.49 | 0.55 | 0.50 | 0.54 | 0.55 | 0.35 | 0.43 | 0.46 | 0.41 |

[a]accessions representative of the heterotic groups identified are shown in bold face; [b]average pairwise genetic distance across group; [c]minimum pairwise distance within group; [d]maximum pairwise distance within group.

enzymes is thought to preclude the formation of a biased population of fragments derived from highly repetitive DNA sequences in the plant genome (Peacock et al, 1981; Carels et al, 1995). Hence, the employment of methylation-sensitive endonucleases is thought to avoid the generation of a biased restriction fragment population in the plant genome. Taking into account the considerable number of markers used in the current investigation, it is impossible to exclude some redundancy due to genetic linkage. However, this source of error is likely to be small in comparison to the large number of polymorphisms identified.

Molecular clustering of the inbred lines considered herein revealed three major heterotic categories (Figure 3). Groups of BSSS-related (I in Figure 3) and LSC-related (II in Figure 3) heterotic material, that track back to the most widely exploited inbreds in temperate regions could be clearly identified as well as a third grouping, formed by more miscellaneous heterotic material, including inbreds developed from crosses between the two previous major heterotic groups, between adapted and exotic germplasm, or derived from distantly related materials (Nelson et al, 2008). In practical breeding programs new lines are often developed from commercial hybrids, i.e. from crosses between heterotic pools. This is not surprising because proprietary hybrids exploit maximum heterosis to be commercially successful. Most of the inbreds present in the third cluster are almost exclusively derived from BSSS to LSC intercrosses. Several inbred lines, such as Lo1242, Lo1322, Lo1297, and Lo1265 were clustered in this miscellaneous category. Although these Lo lines were selected to have good combining ability with lines of the BSSS and LSC groups, the genetic background of both these heterotic germplasms is very likely present in these lines, because the predominant basic genotypes used at the Bergamo maize station for the extraction of breeding materials for the constitution of superior hybrids, adapted to cultivation in the Po valley, the main area of maize production in Italy, were mostly derived from US Corn Belt and Flint complexes. Moreover, recycling of elite inbred lines by two parent crosses and back-crosses was the prevalent method

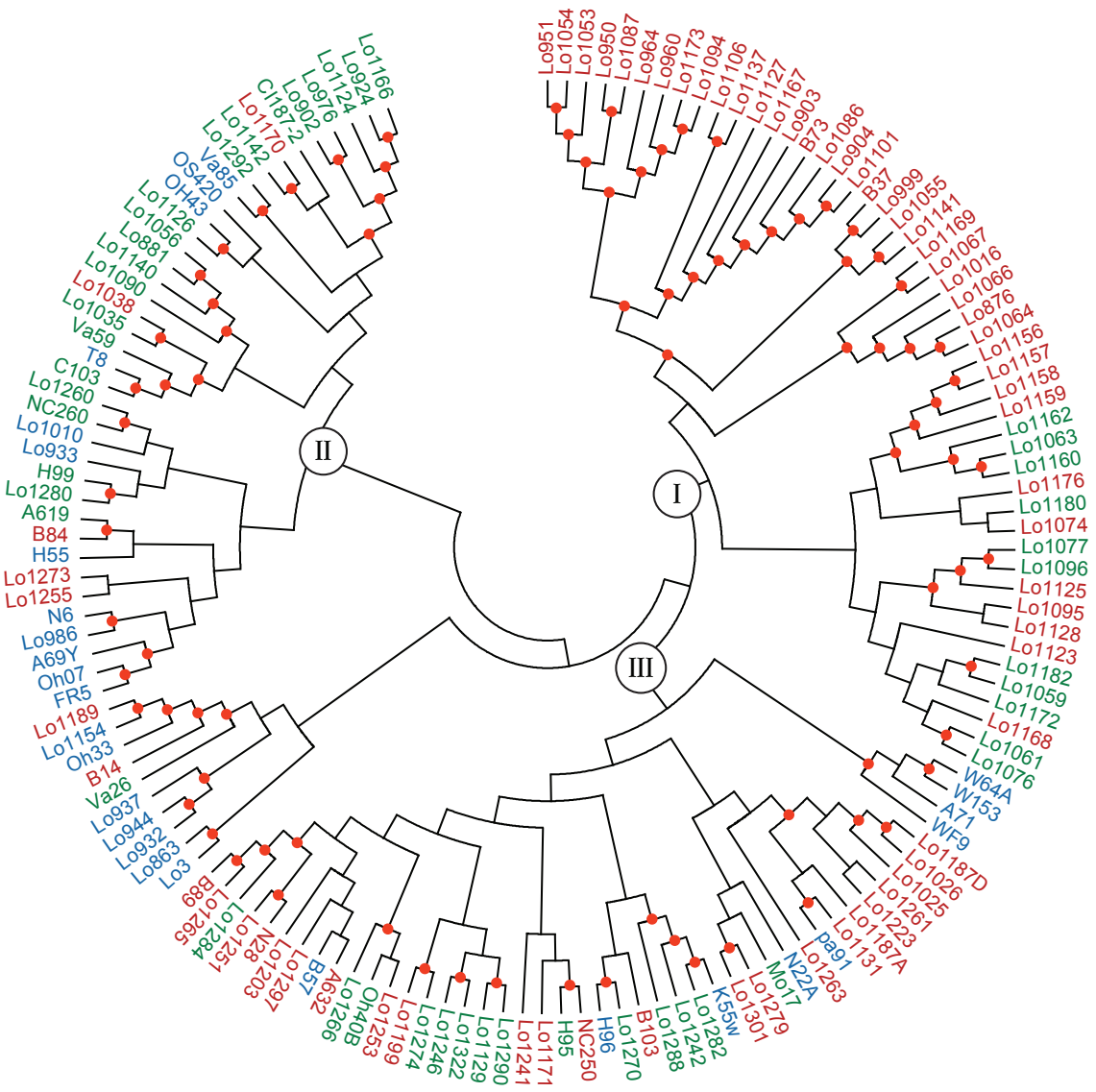


Figure 3 - Neighbor-joining bootstrap clustering of 144 Italian and refernce maize inbred lines based on AFLP markers. Branches, significant at a 67% bootstrap cut-off value, are indicated with red dots. The three major heterotic groups are indicated with roman numbers. Inbreds are color coded following their assumed background as reported in Tables 1 and 2 (red = BSSS; green = LSC; blue = miscellaneous).

used by the Bergamo maize breeding station during the devlopment of the lines anlyzed in this study (Bertolini et al, 1991, 2000).

In general, the observed grouping agreed with available pedigree information even though some discrepancies were noted. These may arise because pedigree relationships are based on identity by descent, whereas the relationships in the phenogram reflect the presence of DNA sequences that are alike in state (Falconer, 1981). For example, Mo17 was developed from the cross CI187.2 x C103, the former originated from Krug and the latter from an LSC strain (Stringfield, 1959). For this reason, several workers (e.g. Smith et al, 1985) hesitate to assign Mo17 to the LSC heterotic group, although in crosses with lines from BSSS or Reid Yellow Dent (RYD) Mo17 behaves like a “typical” LSC line. Indeed, our AFLP data indicate that Mo17 is loosely related with its parent C103 at the molecular level, while clustering in the miscellaneous pool with lines related to commercial hybrids (e.g. Lo1288 and Lo1131) and with Pa91, which originated from a cross between RYD and LSC germplasm. This, furthermore, suggests that Pa91 inherited a larger proportion of its genome from RYD than expected on the basis of its pedigree. Similarly, the BSSS-related B14A line was found in the miscella-

Table 5 - Molecular Analysis of Variance.

| | between groups | between populations within groups | within populations | | | |
|---|---|---|---|---|---|---|
| populations | V % | V % | V % | $F_{ST}$ | $F_{SC}$ | $F_{CT}$ |
| large | 5.93 | 43.42 | 50.65 | 0.493 | 0.461 | 0.059 |
| small | 6.42 | 43.91 | 49.67 | 0.503 | 0.469 | 0.064 |

neous group in association with lines having HY, one of the 16 progenitors of the BSSS population, as their predominant ancestor (Hallauer et al, 1983), while merging more distantly with A632, a B14 related line.

According to Mumm and Dudley (1994) discrepancies may arise due to the fact that clusters obtained with UPGMA are not-overlapping, i.e. an inbred resulting from the cross of two lines also included in the study, can be grouped with only one of the parents. Therefore, the grouping in the phenogram is somewhat artificial in that assignments indicate the group with which the inbred is most similar rather than all similar groups. Despite the discrepancies noted, cluster analysis broadly agreed with pedigree information. In addition, the grouping obtained by cluster analysis was supported by PCA, lending credibility to the classification.

In Italy, maize breeders have relied on the maintenance and exploitation of two or more heterotic breeding groups for the development of superior hybrids. As stated by Hallauer et al (1998), the currently dominant heterotic groups neither are the result of systematic breeding efforts nor are they clearly defined. Our results indicated that the genetic pools can be resolved in in no more than 10 groups of related inbred lines. In this context, Mikel and Dudley (2006) have reported that much of today's materials for hybrid development is derived from approximately seven progenitor lines. Our AMOVA data herein reported suggest that a large proportion of available genetic diversity is found at the within-group level, with the variation present among populations being higher than that among groups of inbreds. In this respect, similar results were obtained in different studies on maize crop varieties using molecular markers (e.g. Vaz Patto et al, 2004 and references therein; Tommasini et al, 2003). In fact, the maize genome exhibits an extraordinarily high level of genetic diversity among inbred lines as assayed at the level of single nucleotide polymorphisms, InDel polymorphism, and structural variation (e.g. Gore et al, 2009; Springer et al, 2009), which are believed to contribute to substain the phenotypic diversity and plasticity of this crop.

The large proportion of variation residing at the within group level suggests that there would be enough variation at this level to select parents to generate new synthetic populations. This could lead to the development of well-characterized pools to select parents contributing good adaptation, persistence and yield. In the long term though, and to avoid exhausting the variability existing at the within group level, it would be advisable to monitor the levels of genetic diversity available and to introgress valuable alleles from other genetic resources, to prevent the loss of complementary gene interactions.

In conclusion, the results presented here indicate that AFLPs are useful in assigning maize inbred lines to heterotic groups and in assessing pedigree relationships among inbred lines. The assignment of maize inbreds to heterotic groups before field testing may allow the breeder to curtail costs by avoiding crosses between groups. Moreover, it should be possible to select new sources for line development that have good chances of yielding superior lines in order to maximize heterosis and consequently yield performance.

## Acknowledgements

This research was financed: Ministero per le Politiche Agricole, Alimentari e Forestali - project Interventi per contrastare la diffusione e I danni da diabrotica nella maiscoltura Italiana (IDIAM).

## References

Ajmone-Marsan P, Gorni C, Chitto' A, Redaelli R, van Vijk R, Stam P, Motto M, 2001. Identification of QTLs for grain yield and grain-related traits of maize (*Zea mays* L.) using an AFLP map, different testers, and cofactor analysis. Theor Appl Genet 102: 230-243

Bertolini M, Bosio M, Bressan M, Coppolino F, Di Fonzo N, Gentinetta E, Introzzi F, Lupotto E, Maggiore T, Perenzini M, Snidaro M, Valoti P, Verderio A, Bianchi A, Soave C, Motto M, Salamini F, 1991. Breeding activity of the maize Station of Bergamo. Synthetic gene pools and inbreds released in the period 1975-1989. Maydica 36: 87-106

Bertolini M, Mazzinelli G, Sartori G, Chittò A, Verderio A, Introzzi F, Bressan M, Pino S, Lorenzoni C, Bonardi P, Motto M, 2000. Breeding activity of the maize Station of Bergamo: inbreds released in the period 1989- 1998. Maydica 45: 73-87

Bonin A, Ehrich D, Manel S, 2007. Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. Mol Ecol 16: 3737–3758

Buckler ES, Gaut BS, McMullen MD, 2006. Molecular

and functional diversity of maize. Curr Opin Plant Biol 9: 172-176
Carels N, Barakat A, Bernardi G, 1995. The gene distribution of the maize genome. Proc Natl Acad Sci USA 92: 11057-11060
Chittò A, Bertolini M, Hartings H, Verderio A, Motto M, 2000. AFLP-based genetic relationships among maize inbred lines selected in a climatically temperate location. Maydica, 45: 257-266
Crossa J, Hernandez CM, Bretting P, Eberhar SA, Taba S, 1993. Statistical genetic considerations for maintaining germ plasm collections. Theor Appl Genet 86: 673-678
Duvick DN, Smith JSC, Cooper M, 2004. Changes in performance, parentage, and genetic diversity of successful corn hybrids, 1930–2000. pp. 65-97. In: Corn: Origin, History, Technology, and Production. Smith CW, Betran J, Runge ECA (eds). John Wiley & Sons, New York, USA
Excoffier L, Lischer HEL, 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10: 564-567
Falconer DS, 1981. Introduction to quantitative genetics. Longman, London, UK
Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JA, McMullen MD, Grills GS, Ross-Ibarra J, Ware DH, Buckler ES, 2009. A first-generation haplotype map of maize. Sience 326: 1115-1116
Hallauer AR, Russell WA, Lamkey KR, 1998. Corn breeding. pp. 463-564. In: Sprague, GF, Dudley, JW (eds). Corn and corn improvement, Third edition
Hammer Ø, Harper D, Ryan PD, 2001. PAST: Paleontological statistics software package for education and data analysis. Palaeontol Electronica 4 art 4
Karp A, Kresovich S, Bhat KV, Ayad WG, Hodgkin T, 1997. Molecular tools in plant genetic resources conservation: a guide to the technologies. IPGRI Technical Bulletin No. 2 International Plant Genetic Resources Institute, Rome, Italy.
Law JR, Koebner RMD, Donini P, Reevers JC, Cooke RJ, 1998. DNA profiling and plant variety registration. The statistical assessment of distinctness in wheat using amplified fragment length polymorphisms. Euphytica 102: 335-342
Messing E, Dooner H, 2006. Organization and variability of the maize genome. Curr Opin Plant Biol 9: 157-163
Mikel MA, Dudley JW, 2006. Evolution of North American dent corn from public to proprietary germplasm. Crop Sci 46: 1193-1205
Mumm RH, Dudley JW, 1994. A classification of 148 U.S. maize inbreds. I: Cluster analysis based on RFLPs. Crop Sci 34: 842-851
Nelson PT, Goodman MM, 2008. Evaluation of Elite Exotic Maize Inbreds for Use in Temperate Breeding. Crop Sci 48: 85-92
Nei M, Li WH, 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76: 5269-5273
Peacock WJ, Dennis ES, Rhoades MM, Pryor AJ, 1981. Highly repeated DNA sequence limited to knob heterochromatin in maize. Proc Natl Acad Sci USA 78: 4490-4494
Pollack LM, 2003. The history and success of the public-private project on germplasm enhancement of maize (GEM). Adv Agron 78: 45-87
Reif JC, Hamrit S, Heckenberger M, 2005. Trends in genetic diversity among European maize cultivars and their parental components during the past 50 years. Theor Appl Genet 111: 838–845
Rohlf FJ, 1993. NTSYS-PC numerical taxonomy and multivariate analysis system. Version 1.8. State Univ, New York, USA
Smith JSC, Goodman MM, Stuber CW, 1985. Genetic variability within US maize germplasm. I. Historically important lines. Crop Sci 25: 550-555
Springer NM, Ying K, Fu Y, Ji T, Yeh C-T, Jia Yen, Wu W, RIchmond T, Kitzman J, Rodenbaum H, Iniguez AL, barbazuk WB, Jeddeloh JA, Nettleton D, Schnable P, 2009. Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genet 5: e1000734
Stich B, Maurer HP, Melchinger AE, Frisch M, Heckenberger M, Rouppe van der Voort J, Peleman J, Sørensen AP, Reif JC, 2006. Comparison of Linkage Disequilibrium in Elite European Maize Inbred Lines using AFLP and SSR Markers. Mol Breed 17: 217-226
Stringfield GH, 1959. Maize inbred lines of Ohio. Ohio Agricultural Experiment Station.
Tommasini L, Batley J, Arnold G, Cooke R, Donini P, Lee D, Law J, Lowe C, Moule C, Trick M, 2003. The development of multiplex simple sequence repeat (SSR) markers to complement distinctness, uniformity and stability testing of rape (*Brassica napus* L.) varieties. Theor Appl Genet 106: 1091-1101
Vaz Patto MC, Satovic Z, Pêgo S, Fevereiro P, 2004. Assessing the genetic diversity of Portuguese maize germplasm using microsatellite markers. Euphytica 137: 63-72
Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Fritters A, Pot J, Peleman J, Kuiper M, Zabeau M, 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23: 4407-4414
Weir BS, Cockerham CC, 1984. Estimating F-statistics for the analysis of population structure. Evolution 38: 1358-1370
White PJ, Johnson LA, 2003. Corn: chemistry and technology. American Association of Cereal Chemists, USA
Xu Y, 2009. Molecular Plant Breeding. CAB International, Wallington, UK