Theor Appl Genet (2008) 117:831–842
DOI 10.1007/s00122-008-0823-2

ORIGINAL PAPER

# Assessment of genetic diversity and relationships among maize (*Zea mays* L.) Italian landraces by morphological traits and AFLP profiling

**H. Hartings · N. Berardo · G. F. Mazzinelli · P. Valoti · A. Verderio · M. Motto**

Received: 24 August 2007 / Accepted: 8 June 2008 / Published online: 27 June 2008


**Abstract** In the present study we have analyzed the genetic diversity pattern in a sample of 54 Italian maize landraces, using morphological traits and molecular markers. Although the 54 landraces surveyed in this study were restricted to Lombardy, the core region of maize production in Italy, our data revealed a large genetic heterogeneity for both morphological and molecular traits in the accessions analyzed. Additionally, our data confirm that the AFLP markers produced a high frequency of polymorphic bands and were able to unequivocally fingerprint each of the landraces considered. Cluster analysis based on AFLP markers displayed a clearer separation of the accessions in comparison to morphological data. Different populations were divided into four major clusters reflecting the geographical origin and seasonal employment of the landraces analyzed. Molecular analysis of variance showed significant ($P < 0.01$) differences among groups, among populations within groups, and among individuals within populations. Approximately 74% of the total variance could be attributed to differences within populations. Conversely, a lower level of differentiation was detected among groups ($\sim$4%). Regarding population structures, the genetic distance between populations ($F_{ST} = 0.25 \pm 0.3$) and the degree of inbreeding within groups ($F_{SC} = 0.22 \pm 0.2$), did not diverge significantly, while both significantly differed from the degree of relatedness between markers within groups ($F_{CT} = 0.04 \pm 0.03$). Results are discussed in relation to a suitable conservation method.

Communicated by A. Charcosset.

H. Hartings · N. Berardo · G. F. Mazzinelli · P. Valoti · A. Verderio · M. Motto (✉)
CRA-Istituto Sperimentale per la Cerealicoltura, Sezione di Bergamo, Via Stezzano 24, 24126 Bergamo, Italy
e-mail: mario.motto@entecra.it

H. Hartings
e-mail: hans.hartings@entecera.it

## Introduction

Because of strong commercial pressures, breeders are forced to draw from a narrow range of tested elite germplasms during the development of modern hybrids. Although the substitution of landraces by elite cultivars is generally admitted to cause a loss of diversity (reviewed in Pollack 2003), it has still to be proven that plant breeding inevitably leads to the loss of genetic diversity. In this context, previous studies have shown that breeding mainly leads to qualitative rather than quantitative shifts in registered cultivars (Le Clerc et al. 2006 and references therein). Thereby, maize breeders have become more aware of the needs of maintaining genetic diversity among hybrid varieties and improving the management of genetic resources through the conservation of traditional populations, that is, landraces. These are populations with high genetic variability and fitness to the natural and anthropological environments where they have originated (Brandolini 1969). They represent not only a valuable source of potentially useful traits, such as resistance or tolerance to biotic and environmental stresses, but also an irreplaceable bank of highly co-adapted genotypes. In this respect, Rebourg et al. (2003), in a study on the introduction and dispersion of American maize in Europe, have found that introductions of Northern American flint populations have played a key role in the adaptation of maize to

the European climate by contributing alleles favoring maize cultivation in temperate environments.

The analysis of the amount and distribution of genetic variation within and among plant populations of a crop species can increase the understanding of the historical process underlying the genetic diversity and provide basic information for breeding programmes and for the establishment of projects to conserve genetic resources. Achieving this goal in curating gene banks is hampered by rising costs, decreasing budgets, and large collection sizes. Therefore, germplasm collection sizes should be optimized to provide both a maximal preservation of genetic variation and a minimal redundancy with regard to genotypes, gene complexes, or possibly even genes (Kresovich et al. 1992).

A broad approach using phenotypic and molecular markers is required to analyze diversity and to support conservation, management, and development of plant genetic resources (Hammer et al. 1999). Phenotypic markers have been of great value in studies of crop evolution (Gepts 1993), germplasm evaluation (e.g., Bretting and Wildrlechner 1995), and in revealing differences between varieties (Gilliland et al. 2000). Molecular markers provide a direct measure of genetic diversity and go beyond indirect diversity measures, based on morphological traits or geographical origin. They have been successfully employed to characterize genetic diversity and for the identification of closely related genotypes (Métais et al. 2002, and references therein). Both kinds of markers have their advantages and drawbacks and their combined utilization is recommended to increase the resolving power of genetic diversity analyses (Singh et al. 1991).

Among maize landraces, genetic variability has been primarily characterized by using morphological traits (Goodman and Bird 1977, and references therein) and isozymes (e.g., Revilla et al. 1998, and references therein). Furthermore, in this crop molecular DNA markers have proven an excellent tool for the assessment of genetic relationships and to monitor genetic diversity and relatedness among elite breeding lines (Smith et al. 1997; Ajmone Marsan et al. 1998; and references therein) and landraces (Rebourg et al. 2003; Reif et al. 2005, and references therein). The amplified fragment length polymorphism (AFLP; Vos et al. 1995) and other multilocus techniques are very useful for the analysis of within-species variation because they allow the rapid acquisition of a large amount of genetic information due to their capability to simultaneously identify a large number of amplification products, that is, a high-multiplex ratio (reviewed in Bonin et al. 2007). The AFLP technique has been largely used in maize to construct genetic maps or to study phylogenic relationships, genetic diversity and other applications (e.g., Lübberstedt et al. 2000; Ajmone Marsan et al. 2001, and references therein).

The objectives of this study were to evaluate the genetic diversity and relationship of 54 accessions of Italian landraces collected in the 1950s in Lombardy, immediately after hybrid introduction. Genetic diversity and relatedness were analyzed with morphological traits and with AFLP markers using single plant DNA samples. As found in previous works, the analysis of individuals has the advantage of estimating the heterogeneity within varieties and allows a more reliable scoring than the analysis of bulks (e.g., Tommasini et al. 2003, and references therein). Information for an appropriate conservation and management of maize germplasm is given.

## Materials and methods

### Plant materials

Fifty-four Italian maize landraces (Table 1) cultivated up to 1950s in Lombardy were considered in this study. The collection sites are distributed across the area of interest (Fig. 1) covering different geographical territories. These were extracted from the germplasm collection of the CRA-Maize Research Unit, Bergamo (Italy). Each population has been regenerated periodically by bulk intercrossing at least 200 plants. All entries were grown in field trials at Bergamo in 2003, 2004, and 2005, in a randomized complete block design with three replications. The experimental plots consisted of four-rows containing 25 plants per row. The density was 57,000 plants per ha. Recommended crop-management techniques were applied. Irrigation was applied regularly as needed.

For each plot, 20 morphological traits (Table 2) were measured from 10 competitive plants. The traits included regarded plant architecture, as well as ear, tassel and kernel related traits. The flowering dates (pollen shedding and silking) were converted into growing degree days (GDD) as:

$$\mathrm{GDD} = \sum_{d=1}^{n} [(TX_d + TN_d)/2 - Tb],$$

where $n$ is the number of days from planting to flowering, $TX_d$ and $TN_d$ are, respectively, the maximum and minimum temperatures (°C) of day $d$, $Tb$ (6°C) is the critical temperature under which development is arrested. Kernel-related traits included were test weight, specific weight, volume, percentages of total protein, starch, oil, and a milling index, that is, the predicted percentage of corn grits in dry-milling, determined using near infrared spectroscopy (NIR) following the development of calibration curves for each traits after chemical analysis as described (Motto et al. 1978). All morphological data are available on request.

**Table 1** Origin of the 54 Italian maize landraces used in this study

| Identification | Denomination | Collection Site |
|---|---|---|
| VA33 | Locale Fiorine | Clusone (BG[a]) (BG) |
| VA34 | Bani | Erbusco (BS) |
| VA35 | Quarantino | Erbusco (BS) |
| VA36 | Nostrano | Erbusco (BS) |
| VA37w | Quarantino Bianco | Erbusco (BS) |
| VA38 | Quarantino Nostrano | Dello (BS) |
| VA39 | Quarantino Nostrano | Buffalora (BS) |
| VA40 | Quarantino Nostrano | Brescia |
| VA41 | Quarantino Nostrano | Paderno Franciacorta (BS) |
| VA42 | Cinquantino | Fenegrò (CO) |
| VA43 | Brianzolo | Garbagnate Monastero (VA) |
| VA44 | Taiolone | Stagno Lombardo (CR) |
| VA45 | Ottofile Mantovano | Motta Baluffi (CR) |
| VA46 | Quarantino S. Famiglia | Stagno Lombardo (CR) |
| VA47 | Centino | Stagno Lombardo (CR) |
| VA48 | Quarantino Giallo | Gaidella di Quistello (MN) |
| VA49w | Cinquantino Bianco | S. Benedetto Po (MN) |
| VA50 | Locale di Passirana | Passirana (MI) |
| VA51 | Locale di Rho | Rho (MI) |
| VA52 | Agostinin | Lacchiarella (MI) |
| VA53 | Ottofile | Rozzano (MI) |
| VA54 | Agostinello | Isola Melzese (MI) |
| VA55 | Melgonin | Motta Visconti (MI) |
| VA56 | Marano | Oreno di Vimercate (MI) |
| VA57 | Nostrano dell'Isola | Arcore (MI) |
| VA58 | Scagliolo | Trezzo d'Adda (MI) |
| VA59 | Giallo Agostanello | Concorezzo (MI) |
| VA60 | Giallo Agostano | Concorezzo (MI) |
| VA61 | Ottofile | Zinasco (PV) |
| VA62 | Nostrano dell'Isola | Pala (SO) |
| VA63 | Nostrano Locale | Pala (SO) |
| VA64 | Nostrano Locale | Pala (SO) |
| VA65 | Locale Chiavenna | Verceia (SO) |
| VA66 | Locale Verceia | Verceia (SO) |
| VA67 | Locale Tirano | Barbone (SO) |
| VA68 | Nostrale | Madonna del Piano (SO) |
| VA69 | Locale | Forte (SO) |
| VA70 | Locale | Somaggia (SO) |
| VA71 | Agostanello | Lonate Pozzolo (VA) |
| VA72 | Nostrano Locale | Besnate (VA) |
| VA73 | Agostanello Locale | Origgio (VA) |
| VA518 | San Pancrazio | Salvagna (BG) |
| VA553 | Scagliolo Marne | Salvagna (BG) |
| VA558 | Rostrato | Cantello (VA) |
| VA561 | Locale Rostrato | Fontanella sul Monte (BG) |
| VA569 | Sacra Famiglia | Salvagna (BG) |
| VA571 | Sintetico Zanchi | Salvagna (BG) |
| VA572 | Nostrano dell'Isola Finardi | Salvagna (BG) |
| VA578 | Rostrato | Torre Boldone (BG) |
| VA888 | Cinquantino di Stezzano | Stezzano (BG) |
| VA903 | Cinquantino 2° raccolto | Alto Milanese (MI) |
| VA904 | Cinquantino 2° raccolto | Alto Milanese (MI) |
| VA1196 | Rostrato di Valchiavenna | Chiavenna (SO) |
| VA1210 | Rostrato | Carenno (LC) |

[a] Province of collection site: *BG* Bergamo, *BS* Brescia, *CO* Como, *CR* Cremona, *LC* Lecco, *MI* Milano, *MN* Mantova, *PV* Pavia, *SO* Sondrio, *VA* Varese

## Molecular analyses

For each population in Table 1, we extracted DNA from thirty individual shoots of 2-week-old germinated seedlings as described in Chittò et al. (2000). Sample sizes were chosen according to Crossa et al. (1993) and Krauss (2000), who found that most procedures for estimating diversity based on AFLP data yield accurate results when approximately 30 individuals are analyzed per population. Molecular genotyping was carried out using the AFLP protocol according to Vos et al. (1995). Briefly, total DNA of each sample was double-digested with *Eco*RI and *Mse*I restriction enzymes and specific adaptors were ligated onto the digested DNA, pre-amplification was conducted with *Eco*RI primer (E00 + A as a selective nucleotide). Upon pre-amplification, selective amplification was conducted using 3 selective nucleotides for both *Eco*RI and *Mse*I primers. The base sequences of the *Eco*RI and *Mse*I primers together with the specific extensions of the selective primers used are shown in Table 3. Primer combinations with a high polymorphic detection rate in maize were employed (Chittò et al. 2000).

Autoradiographs were manually scored for each major polymorphic amplified fragment. Bands of low intensity were ignored. A two-dimensional data matrix was constructed, reflecting the absence/presence of each polymorphic fragment identified within the landraces analyzed.

## Statistical analyses

Standard analyses of variance were performed on each morphological data to test the significance of variation between accessions. These analyses allowed us to estimate genotypic and environmental variances, as well as broad sense heritability (per-landrace-mean basis) for each trait.

The 20 morphological descriptors (Table 2) were used to (1) derive Mahalanobis distances and (2) in principal

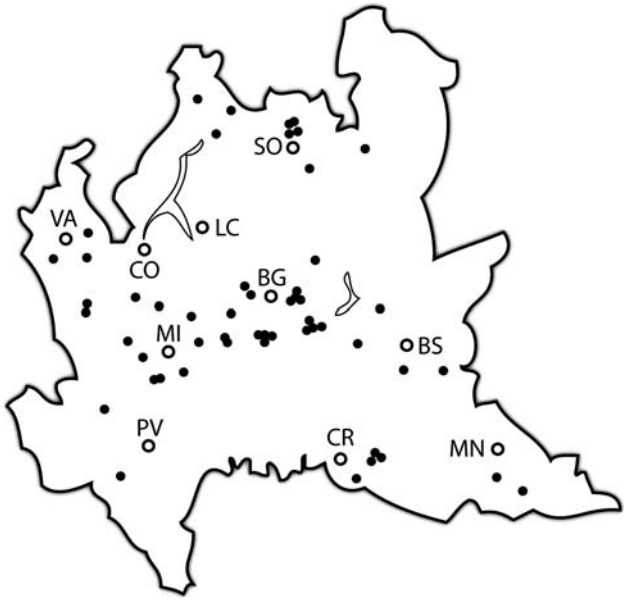


**Fig. 1** Schematic map of Lombardy. The position of capital cities is indicated with *open circles*. Collection sites are indicated with *dots*. Abbreviations are as indicated in Table 1

component analysis (PCA) allowing to determine the traits most effective in discriminating between landraces. Common components coefficients, eigenvector values, and relative and cumulative proportions of the total variance expressed by single traits were calculated. Traits with a correlation >0.6 were considered as relevant for that component. In addition, a dissimilarity matrix based on Euclidean distance coefficients was generated to assess the level of dissimilarity between landraces. All calculations and analysis were made using the appropriate options of NTSYS-pc version 1.80 (Rohlf 1993).

The AFLP data were evaluated by means of polymorphism information content (PIC) of a marker (Smith et al. 1997). This value is defined as the probability that two alleles taken at random from a population can be distinguished using the marker in question, is a measure of allele diversity at a locus. The PIC was calculated by the formula $PIC = 1 - \Sigma f_i^2$, where $f_i$ is the frequency of the ith allele. Because the AFLP technology produces dominant markers, only two states can be observed at each band position and consequently, the highest PIC value that can be obtained equals 0.5.

The genetic similarities (GS) were calculated from AFLP data according to Nei and Li (1979) as $GS_{ij} = 2N_{ij}/(N_i + N_j)$, 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, considering all primer combinations. GS values were converted into genetic distances: GD = 1 – GS, which

**Table 2** List of the 20 morphological descriptors utilized in this study with variety adjusted means and associated standard deviations, ranges, and broad sense heritabilities computed on a population mean basis as obtained for the landraces used

| Abbreviation | Morphological descriptor | Mean | SD | Range | | $h^2$ |
|---|---|---|---|---|---|---|
| | | | | Min | Max | |
| LWT | Leaf width (cm) | 9.5 | 1.44 | 6.0 | 12.0 | 0.78 |
| PHT | Plant height (cm) | 166.0 | 27.43 | 110.0 | 215.0 | 0.72 |
| EHT | Ear height (cm) | 77.0 | 22.42 | 32.0 | 120.0 | 0.82 |
| GDDP | Growing degree days pollen shed | 682.3 | 76.52 | 528.0 | 801.0 | 0.78 |
| GDDS | Growing degree days silking | 710.5 | 75.10 | 551.0 | 825.0 | 0.81 |
| PBN | Primary branch number (no) | 13.2 | 2.90 | 6.0 | 20.0 | 0.72 |
| BTL | Branched tassel length (cm) | 42.4 | 6.10 | 31.0 | 60.0 | 0.61 |
| TLG | Tassel length (cm) | 20.2 | 3.42 | 13.0 | 28.0 | 0.65 |
| ELG | Ear length (cm) | 16.9 | 2.87 | 12.0 | 24.0 | 0.77 |
| EDI | Ear diameter (cm) | 40.1 | 4.74 | 31.0 | 50.0 | 0.85 |
| CDI | Cob diameter (cm) | 25.9 | 3.53 | 19.0 | 33.0 | 0.68 |
| ERN | Ear row number (no) | 13.1 | 2.42 | 8.0 | 20.0 | 0.88 |
| KWE | Kernel weight ($g \times 10^3$) | 245.0 | 56.46 | 155.0 | 420.0 | 0.77 |
| KTW | Kernel test weight (kg/hl) | 77.2 | 3.75 | 67.2 | 83.8 | 0.81 |
| SKW | Specific kernel weight ($g/cm^3$) | 1.3 | 0.04 | 1.2 | 1.4 | 0.79 |
| KRV | Kernel volume ($cm^3$) | 19.7 | 5.09 | 12.0 | 33.0 | 0.83 |
| MLI | Milling index (%) | 53.6 | 2.54 | 45.9 | 57.1 | 0.67 |
| KOC | Kernel oil content (%) | 6.0 | 1.24 | 3.5 | 8.3 | 0.63 |
| KSC | Kernel starch content (%) | 65.3 | 3.06 | 58.6 | 73.1 | 0.60 |
| KPC | Kernel protein content (%) | 11.0 | 1.15 | 8.3 | 13.7 | 0.64 |

**Table 3** List of AFLP primers used with *Eco*RI and *Mse*I base sequences (left columns), primer codings (center columns) and specific trinucleotide primer extensions (right columns)

| EcoRI | | | MseI | | |
|---|---|---|---|---|---|
| GACTGCGTACCAATTC.NNN | E31 | AAA | GATGAGTCCTGAGTAA.NNN | M47 | CAA |
| | E32 | AAC | | M48 | CAC |
| | E33 | AAG | | M49 | CAG |
| | E35 | ACA | | M50 | CAT |
| | E36 | ACC | | M51 | CCA |
| | E38 | ACT | | M60 | CTC |
| | | | | M61 | CTG |

were used for cluster analysis with NTSYS-pc software producing a dendrogram using the unweighted pair-group arithmetic method (UPGMA).

The significance of clustering was assessed by means of bootstrap analysis with PAST software (Hammer et al. 2001) applying 1,000 replicates. Bootstrap values exceeding a 50% cut-off are indicated above the corresponding clusters in the respective figures.

An analysis of molecular variance (AMOVA) was performed on AFLP data to partition the total genetic variation across the landraces. AMOVA and significance tests were performed using Arlequin version 3.1 (Excoffier et al. 2005). The degree of inbreeding within groups ($F_{SC}$), the degree of relatedness between markers within groups ($F_{CT}$), as well the fixation index ($F_{ST}$) were computed according to Weir and Cockerham (1984). $F_{ST}$ measures the genetic effect of landrace subdivision as the proportional reduction in overall heterozygosity owing to variation in allele frequencies among different landrace populations. Values of $F_{ST}$ range from 0 (completely undifferentiated) to 1 (completely differentiated). Populations with little divergence have $F_{ST}$ values less than 0.05, while moderately differentiated populations have $F_{ST}$ values between 0.05 and 0.15, greatly differentiated populations have $F_{ST}$ values between 0.15 and 0.25, and very greatly differentiated populations have $F_{ST}$ values greater than 0.25 (Hartl and Clark 1997). Gene flow was estimated from $F_{ST}$ as follows: $N_m = 0.25(1 - F_{ST})/F_{ST}$ (Nei and Li 1979). The result is independent of population size because the force of gene flow, which is measured by the fraction of migrants in a population (denoted as m), is counteracted by the force of genetic drift, which is proportional to the inverse of the population size ($N$). $N_m < 1$ indicates a local differentiation of populations, while $N_m > 1$ when a little differentiation among populations occurs (McDermott and McDonald 1993).

## Results

### Morphological variation

The variety adjusted mean, range and standard deviation of the 20 phenotypic traits examined, are reported in Table 2. An estimation of heritability, made on the assumption that the accessions taken into consideration are representative of a reference population, is also provided in this table. The Italian landraces considered displayed ample variation for earliness, plant architecture traits, tassel traits, and ear and kernel characteristics. Plant height varied from 110 cm for the landraces grown in unirrigated areas to 215 cm for those from irrigated plains. Leaf width ranged from 6 to 12, while GDD for female anthesis was between 551 and 825. Ear type varied from true cylindrical to extra conical, with a row number varying from 8 to 20 and ear diameter ranging from 31 to 50 mm. Average ear length varied from 12 to 24 cm, while kernel shape and size were also highly variable. Protein concentrations between 8.3 and 13.7% were found, while starch ranged from 58.6 to 73.1%, and oil from 3.5 to 8.3%. Average kernel test weight and specific kernel weight ranged from 67.2 to 83.8 kg/hl and from 1.15 to 1.38 g/cm$^3$, respectively. Finally, the milling index ranged from 45.9 to 57.1%. High heritability values were observed for the majority of the traits.

Relationships between traits were investigated using graphs, correlation coefficient estimations and PCA. The first five principal components (PCs) accounted for 86.3% of total variation (Table 4). In the first PC, which explained 56.1% of the total variance, the most important traits were earliness and correlated traits related to plant, tassel, and kernel size. In the second PC (10.6%), predominant traits were specific kernel weight and milling index, while the third PC (9.3%) described variation in traits related to ear diameter. Correlation of traits is evident when PC's are explored graphically. Plotting of PC1 against PC2, which accumulatively account for $\sim$67% of the total variance, allows visualizing three major correlation groupings. Groups comprising plant morphology traits, seed morphology traits, and seed quality traits could be identified. Protein and starch content showed little relationship with the other quality traits measured (Fig. 2). The first two PCs obtained in PCA analysis on the whole set of 54 landraces characterized are illustrated in Fig. 3. The first PC separated landraces according to precocity, plant and ear height, and kernel weight, while the second PC divided the populations according to kernel volume and specific weight. It is notable that VA61 and VA578 were isolated from the other landraces.

Cluster analysis was used to reveal the association between landraces. Genetic similarity was calculated from the morphologic data by UPGMA cluster analysis based on

**Table 4** Eigenvectors and accumulated variation of the first five components (PC) from the morphological correlation matrix derived from 54 Italian maize landraces

| Descriptor | PC1 | PC2 | PC3 | PC4 | PC5 |
|---|---|---|---|---|---|
| LWT[a] | **−0.83*** | −0.11 | 0.09 | −0.02 | −0.19 |
| PHT | **−0.81** | −0.21 | 0.03 | −0.01 | −0.04 |
| EHT | **−0.76** | −0.21 | 0.11 | −0.29 | −0.15 |
| GDDP | **−0.92** | −0.12 | −0.01 | −0.22 | 0.06 |
| GDDS | **−0.92** | −0.14 | −0.01 | −0.22 | 0.04 |
| PBN | **−0.71** | −0.43 | 0.01 | 0.01 | 0.05 |
| BTL | **−0.77** | −0.09 | −0.15 | 0.11 | 0.02 |
| TLG | **−0.67** | −0.02 | −0.40 | 0.18 | 0.10 |
| ELG | −0.57 | −0.39 | −0.35 | 0.01 | 0.07 |
| EDI | **−0.68** | 0.03 | 0.46 | 0.38 | −0.06 |
| CDI | −0.32 | 0.05 | **0.64** | 0.34 | −0.23 |
| ERN | −0.09 | −0.48 | **0.72** | −0.00 | 0.11 |
| KWE | **−0.61** | 0.53 | −0.26 | 0.39 | −0.15 |
| KTW | 0.23 | **−0.77** | −0.06 | 0.31 | −0.09 |
| SKW | 0.30 | **−0.61** | −0.14 | 0.30 | 0.07 |
| KRV | −0.58 | 0.53 | −0.34 | 0.35 | −0.06 |
| MLI | 0.37 | −0.56 | −0.24 | 0.12 | −0.22 |
| KOC | −0.03 | 0.01 | −0.10 | −0.34 | **−0.84** |
| KSC | −0.15 | 0.47 | 0.48 | −0.01 | 0.06 |
| KPC | 0.52 | −0.05 | 0.06 | 0.28 | −0.40 |
| Total variance (%) | 56.1 | 10.6 | 9.3 | 6.0 | 4.3 |
| Cumulative variance (%) | 56.1 | 66.7 | 76.0 | 82.0 | 86.3 |

[a] For abbreviations see Table 2

* Values >0.6 are presented in bold face and indicate traits important for PC definition

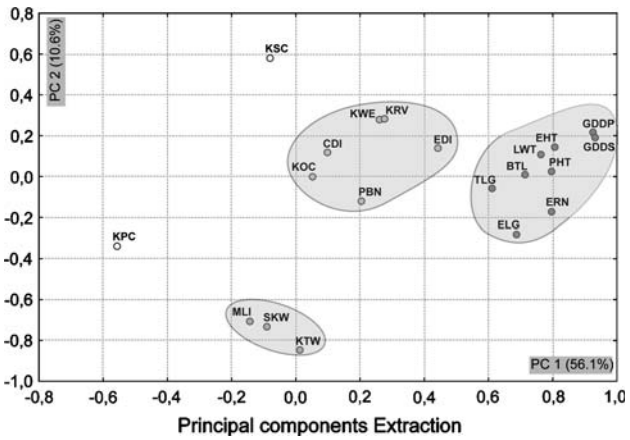


**Fig. 2** Principal component analysis of morphological traits from 54 Italian landraces showing groups of morphological traits analyzed. The first two components are reported. Coding of traits is as indicated in Table 2

Euclidian distance coefficients, with similarity coefficients ranging from 0.07 (VA42 – VA903) to 0.28 (VA47 – VA578). Cluster analysis divided the 54 landraces into four

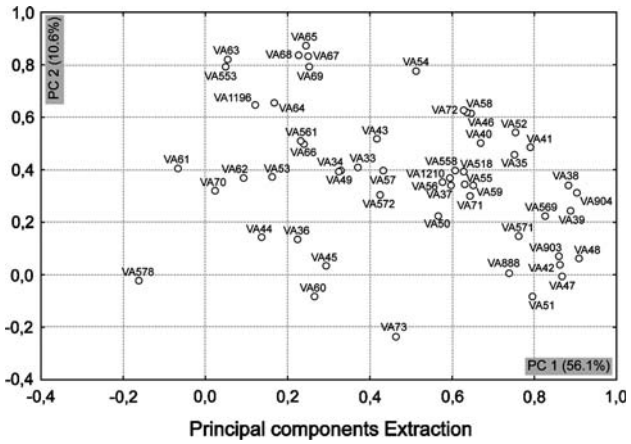


**Fig. 3** Principal component analysis of 54 Italian landraces based on morphological traits

distinctive groups (Fig. 4). A first main cluster (I) included most of the accessions with small kernels, a high number of rows and a short vegetative cycle once used as second harvesting crop. The second main cluster (II) was also highly differentiated. Within this cluster, two distinctive subclusters were clearly identified. A first sub-cluster was based on the *Nostrano* type and derived forms, while in the second sub-cluster most of the varieties from Valtellina, an alpine valley in northern Lombardy were found. The third main cluster (III) was mainly composed of semi-early, eight-row types and derived forms. Main cluster IV was also heterogeneous containing several semi-early *Marano* types.

## Molecular analyses

For each landrace, approximately 100–200 amplified fragments could be visualized in each AFLP run depending on the primer pair employed. In total, the 10 primer pairs used produced 284 polymorphic AFLP bands for the 54 accessions analyzed. Although only major polymorphisms were scored as described, an average of over 28 markers could be obtained for each primer combination, confirming that AFLP analysis is clearly a powerful means of DNA profiling in maize, with substantial polymorphisms between varieties. The number of markers ranged from 12 (primer pair E32M61) to 46 (primer pair E35M48).

The existence of 284 AFLP loci appears sufficient to investigate the genetic structure of the 54 populations, that is, relatively distantly related entities. An average PIC value of 0.37 ± 0.12 was obtained, while individual values ranged from 0.09 to 0.50. Approximately 59% of markers (167 out of 284) used had a PIC exceeding 0.3, demonstrating the good discriminatory power of the markers identified (Fig. 5).

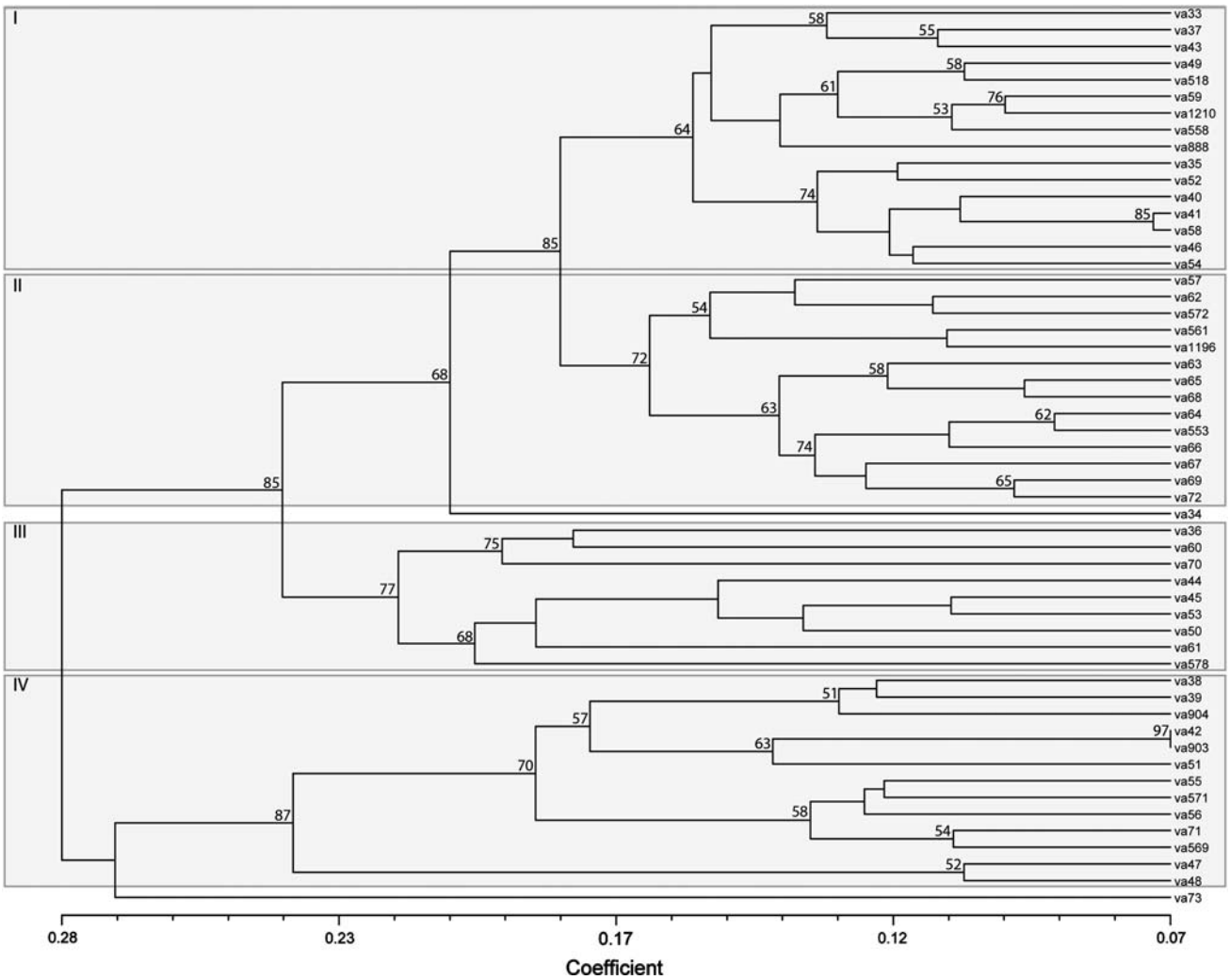


**Fig. 4** UPGMA clustering of 54 Italian maize landraces based on morphological traits. Bootstrap values are indicated

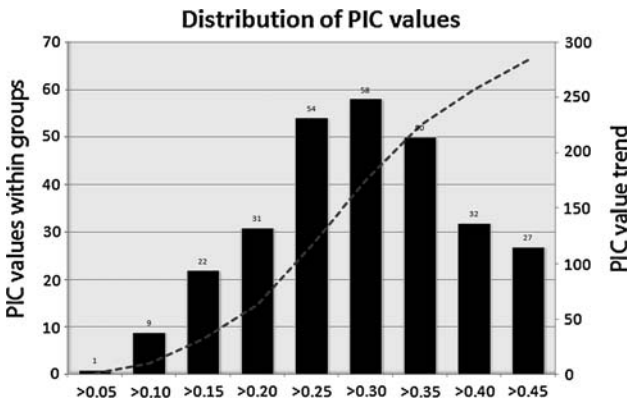


**Fig. 5** PIC values for primer pairs used in AFLP analysis on landraces. PIC values are distributed across nine 5-percentile classes. *Numbers above bars* indicate the amount of PIC values within each class. A *dashed line* indicates the cumulative trend of PIC values

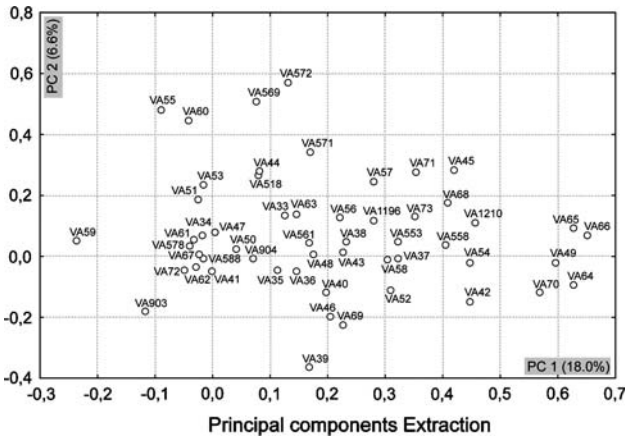


**Fig. 6** Principal component analysis of 54 Italian maize landraces based on AFLP markers

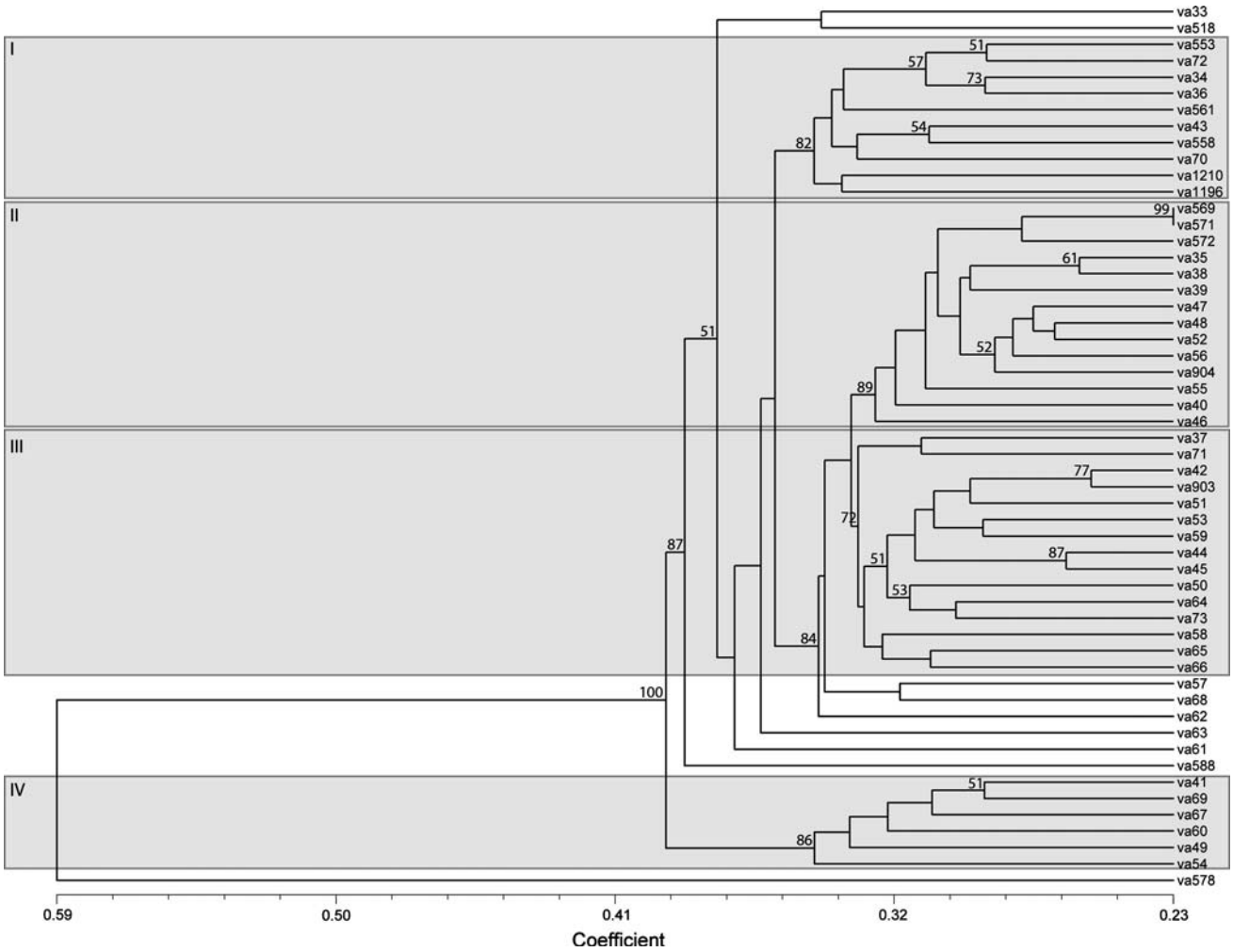


**Fig. 7** Dendrogram based on the genetic similarity index and UPGMA clustering of 54 Italian maize landraces based on AFLP markers

PC analysis based on AFLP markers visualized the distribution of variance within the 54 landraces considered. Figure 6 represents a graphical distribution of the landraces in a plane defined by the first two PCs, which accounted for 24.6% of total variation.

Scoring of the markers allowed the construction of a 284 × 54 binary array, which was subsequently utilized to compute GD values for all pairs of accessions studied. GD values ranged from 0.23 for VA569 and VA571 to 0.59 for the varieties VA50 and VA578. An average GD of 0.437 ± 0.012 was calculated for the entire data set. Distance measures were subsequently used to construct a hierarchical tree using the UPGMA method. In the dendrogram generated from the data set, 4 major clusters could be identified (Fig. 7).

A first heterogeneous cluster (I) included several varieties of *Nostrano* and *Rostrato* derivatives. The second main cluster (II) mainly consisted of varieties with a short vegetative cycle. The third main cluster (III) is highly heterogeneous and contains several varieties of early and semi early eight-row derivatives. This major cluster was further divided in two subclusters: one of these containing representatives of the population with long cylindrical (flint) ears of the *Nostrano dell'Isola* group and from Valtellina. The fourth main cluster was highly differentiated and contained mainly early type accessions that may be seen as a case of particular adaptation.

## Analysis of molecular variance (AMOVA)

To conduct AMOVA (Table 5), the landraces were organized into four main groups as suggested from the dendrogram obtained by grouping landraces based on their AFLP profiles (Fig. 7). Because similar results were obtained with different types of classifications, only data obtained with the first classification are given in Table 5. Highly significant differences were observed among groups, among populations within groups, and among individuals within populations. Approximately 74% of the total variance could be attributed to differences within

**Table 5** Hierarchical analysis of molecular variance (AMOVA) with mean squared deviation (MSD), a summary of *F*-statistics and gene flow ($N_m$)

| Source of variation | MSD | Variance components | % of total variance |
|---|---|---|---|
| Among groups | 1062.1 | 2.1** | 4.5 |
| Among populations | 5200.4 | 10.2** | 21.4 |
| Within populations | 13310.1 | 35.2** | 74.1 |
| *F*-statistics | | | |
| $F_{SC} = 0.22 \pm 0.20$ | | | |
| $F_{ST} = 0.25 \pm 0.30$ | | | |
| $F_{CT} = 0.04 \pm 0.03$ | | | |
| $N_m = 0.75$ | | | |

** Significant at $P < 0.01$

populations. Conversely, a low level of differentiation was detected among groups (~4%). The genetic distance between populations ($F_{ST} = 0.25 \pm 0.3$) did not significantly differ from the extent of inbreeding within groups ($F_{SC} = 0.22 \pm 0.2$). The degree of relatedness between markers within groups ($F_{CT} = 0.04 \pm 0.03$) was significantly low. The sizeable genetic differentiation among landrace populations was confirmed by the gene flow estimate that resulted <1, suggesting a local differentiation of populations.

### Comparison between molecular and morphological distances

Relationship between morphological and molecular distances was significant but low ($r = 0.165$**). Analysis of the graph (Fig. 8) illustrates that this relationship is not linear and that a vast dispersion is present. Low marker distances are associated with low morphological distances. Conversely, high marker distances are associated with both high and low morphological distances. Therefore, marker divergence behaves as a limiting factor of morphological divergence.

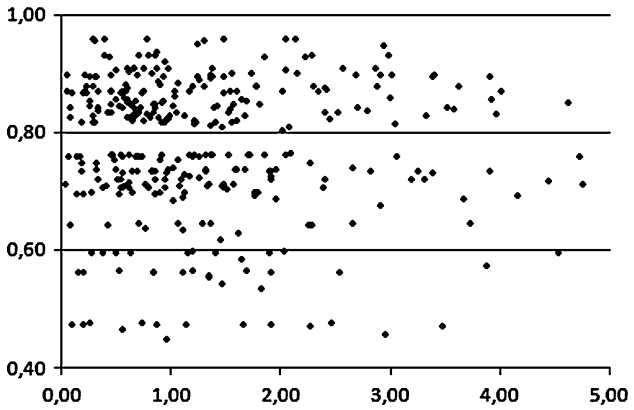


**Fig. 8** Relationships between morphological (*x* axis) and AFLP (*y* axis) distances of 54 Italian maize landraces

## Discussion

The necessity to preserve and characterize genetic maize resources appeared after the introduction of the first commercial hybrids 50 years ago and led to the birth of many collections, both at a joint European and national level, including the Italian collection of the CRA—Maize Research Unit, Bergamo (Brandolini 1969 and references therein). For this latter collection, Brandolini and Brandolini (2001) have recently reported a classification of 562 Italian maize accessions using numerical taxonomic methods of 17 phenological, morphological, and geographical traits. Accessions were grouped into 65 agroecotypes with major clusters containing 34 landraces in nine racial complexes of common ancestors and/or places of origin.

In the present study, we have analyzed the genetic diversity pattern in a sample of 54 Italian landraces, by using both morphological traits and molecular markers. Although the 54 landraces surveyed in this study were restricted to Lombardy, the core region of maize production in Italy, our data revealed a large genetic heterogeneity. The genetic diversity revealed for both morphological and molecular traits indicates that the populations are highly adapted to specific environmental conditions and uses because of centuries of selection and local adaptation and hence could be valuable sources of genetic variability. In fact, since the introduction of maize in Italy, five centuries ago, one might expect that genetic diversity was on one hand enhanced by mutation, meiotic recombination, and new introgression, and on the other curtailed by genetic drift and natural and artificial selection (Hartl and Clark 1997). This would result in numerous open-pollinated varieties adapted to specific local conditions and farmer needs, as is likely the case for the accessions investigated in the current study. Moreover, our results are in good agreement with previous studies documenting that maize is the most diverse crop plant known, containing extensive diversity at both the phenotypic and molecular levels (see Buckler et al. 2006, for a recent review). Additionally, studies on European maize germplasm have already pointed out the presence of a great variability in morphological traits and large molecular diversity in traditional populations (Gauthier et al. 2002; Brandolini and Brandolini 2001 and references therein).

Molecular markers can support a more detailed characterization of genetic resources and were proven powerful tools for studying maize population structures allowing the classification and identification of core subsets (e.g.,

Gouesnard et al. 2005, for a review). A large potential lies in their ability to identify the structure of genetic diversity within and among accessions, which can be relevant for the optimization of collections, the planning of seed regeneration, and the successful implementation of pre-breeding approaches. The results obtained in this study confirm that the use of AFLP markers allowed to unequivocally fingerprint each of the landraces considered and to effectively detect genetic variation in the 54 landraces. In addition, AFLP revealed a large number of polymorphic DNA fragments with an average of 34.7 markers for each primer combination used. An average PIC value of 0.37 was obtained, while approximately 59% of the markers (167 out of 284) used had a PIC exceeding 0.3, demonstrating the good discriminatory power of the markers identified, and suggesting that considerable variation is detectable with AFLP markers. This is in good agreement with previous studies, demonstrating that, in maize, the degree of polymorphism detectable by DNA markers is very high (Smith et al. 1997; Ajmone Marsan et al. 1998, and references therein). Our primer pairs differed in their capacity to amplify bands, but even with the least informative primer pair (E32M61), the number of loci detected was moderately high (12) when compared to several others DNA markers. Each genotype had a unique banding profile for every AFLP primer combination. The results presented here show that AFLPs are able to reveal variability both between and within maize landraces, which has three main purposes for germplasm management: (1) a means of landrace identification, (2) a way to detect genetic diversity, and (3) a means to reveal genetic relationships.

A criticism that is sometime raised against the use of AFLPs for profiling purposes is that the chromosomal locations of the bands are unknown (i.e., they are not mapped). Thus, it is possible that their distribution is unrepresentative of the genome as a whole. To partially overcome this shortcoming, primer combinations were chosen that had previously been utilized in mapping studies and for which, consequently, the genome coverage was known (Castiglioni et al. 1999). Second, AFLP markers are generally scored as dominant markers, thus it is difficult to distinguish heterozygous individuals from homozygotes. Therefore, heterozygosity has to be calculated indirectly, by assuming that gene frequencies in the studied populations are in Hardy–Weinberg equilibrium. Since maize is a highly outcrossing species, this assumption is not widely violated. Taking into account the technical ease of the AFLP technique and its high information content, it is considered an advantageous method for carrying out genetic population studies (Bonin et al. 2007 for a review).

As expected, the genetic variability found in our analysis was ample, with AMOVA suggesting that the largest proportion of available genetic diversity is to be found at the within-population level (74.1%), with the variation present among populations being higher (21.4%) than that among groups of populations (4.5%). Similar results were obtained in different studies on maize (e.g., Vaz Patto et al. 2004, and references therein) and other crop varieties using molecular markers (e.g., Tommasini et al. 2003, and references therein). The large proportion of variation residing at the within population level suggests that there would be enough variation at the population level to select parents to generate new synthetic populations. This could lead to developing well-characterized populations to select parents contributing good adaptation, persistence and yield. In the long term though, and to avoid exhausting the variability existing at the within population level, it would be advisable to monitor the levels of genetic diversity available and to introgress valuable alleles from other populations, to prevent the loss of complementary gene interactions due to inbreeding.

The distribution of genetic diversity within and among populations is a function of the rate of gene flow between populations. The extent of gene flow in a species depends on the distribution of the habitats it occupies, on the size and degree of isolation of its populations, and on the movement of pollen and seeds between populations. In the landraces as a whole, each population proved to be heterogeneous at a large number of loci. This result can be explained considering the high rate of free pollinations within families in any allogamous species such as maize, a low gene flow due to local isolation of the populations, and/or large effective population sizes. Therefore, each sampled landrace is actually a mixture of a large number of distinct genotypes that casually intercross at each generation sharing a common gene pool, which belongs to the landraces because of local adaptation, which may be ascribed to a combination of climatic conditions and agronomic practices.

A low correlation was detected between the observed molecular and morphological variation patterns. This result is in agreement with previous observations reported by a number of workers (Rebourg et al. 2003 and references therein). They showed a tendency towards a triangular pattern (as seen in Fig. 8) in the distribution of data between morphological and molecular distances, supporting the superiority of molecular marker data to define groups of populations with similar origins. Such a triangular pattern might be due to the limited number of molecular and morphological markers examined, to a low gene flow, which would allow for accumulation of adaptative differences, and to the polygenic inheritance of traits. Indeed, Burstin and Charcosset (1997) reported that a given quantitative value can be obtained with different gene combinations and that most of the traits generally considered for phenotypic distance estimation exhibit polygenic

inheritance. Most molecular polymorphisms will be effectively neutral and thus provide no direct assessment of fitness. The forces that cause differentiation for these markers would be the result of mutation, genetic drift, and low gene flow without selection. Conversely, morphological traits are generally believed to be subject to natural selection and their expression is partially under the influence of environmental factors. Moreover, in contrast to morphological traits, AFLP polymorphism is based directly on DNA sequence variation. A change in a nucleotide repetition can result in a change of pattern. Despite these differences, low gene flow would allow accumulation of small or higher adaptive differences, explaining some concordance between morphological and molecular traits. Furthermore, differences in uniformity for different markers might be due to the closeness of some markers to regions of the genome subject to selection, and, therefore, selected indirectly. This hypothesis is supported by previous studies showing that allelic frequencies at some molecular loci are under functional constraints (Boland 1996).

The increasing costs to efficiently manage large ex-situ collections encourage curators to form core subsets and to eliminate residual and duplicated accessions. Although cross subsets can be assembled to facilitate the in-depth study of evolution, and use of genetic resources stored in germplasm collections, this implies a substantial reduction in the number of accessions compared with the initial collection and, thus, a possible reduction in the genetic and phenotypic diversity compared with what exists in the original collection (Hammer et al. 1999). In maize, strategies for the development of core collections were given by a number of workers (see Gouesnard et al. 2005, for a review). The findings of the current study indicate that core subsets can be formed based on morphological or molecular marker data. Overall, our findings suggest that although a high molecular variability can be found among landraces, most plant genotypes belong to the same landraces. The identity of the landraces as a whole seems to have been preserved due to the large number of polymorphism and the presence of specific alleles for local adaptation. From a conservation perspective, the high genetic diversity observed between and within the set of Italian populations suggests to assemble subcore collections by intercrossing the more similar landraces, identified by cluster analysis, and to use an optimal number of individuals from each population in order to preserve most of the less frequent alleles in the subcore collections.

**Acknowledgments** This work was supported by Ministero delle Politiche Agricole, Alimentari e Forestali, Roma, Italy: special grant "Risorse genetiche vegetali-FAO.

## References

Ajmone Marsan P, Castiglioni P, Fusari F, Kuiper M, Motto M (1998) Genetic diversity and its relationship to hybrid performance in maize as revealed by RFLP and AFLP markers. Theor Appl Genet 96:219–227

Ajmone Marsan P, Gorni C, Chittò 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

Boland CR (1996) Setting microsatellites free. Nat Med 2:972–974

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

Brandolini A (1969) European races of maize. Annu Corn Sorghum Res Conf Proc 24:36–48

Brandolini A, Brandolini A (2001) Classification of Italian maize (*Zea mays* L.) germplasm. Plant Gen Res Newslett 126:1–11

Bretting PK, Wildrlechner MP (1995) Genetic markers and plant genetic resource management. In: Janick J (ed) Plant breeding reviews, vol 13. Wiley, New York, pp 11–71

Buckler ES, Gaut BS, McMullen MD (2006) Molecular and functional diversity of maize. Curr Opin Plant Biol 9:172–176

Burstin J, Charcosset A (1997) Relationship between phenotypic and marker distances: theoretical and experimental investigations. Heredity 79:477–483

Castiglioni P, Ajmone-Marsan P, Van Wijk R, Motto M (1999) AFLP markers in a molecular linkage map of maize: codominant scoring and linkage group distribution. Theor Appl Genet 99:425–431

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, Eberhart SA, Taba S (1993) Statistical considerations for maintaining germplasm collections. Theor Appl Genet 86:673–678

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolut Bioinf Online 1:47–50

Gauthier P, Gouesnard B, Dallard J, Redaelli R, Rebourg C, Charcosset A, Boyat A (2002) RFLP diversity within and among European traditional maize populations. Theor Appl Genet 105:91–99

Gepts P (1993) The use of molecular and biochemical markers in crop evolution studies. Evol Biol 27:51–94

Gilliland TJ, Coll R, Calsyn E, de Loose M, van Eijk MJT, Roldan-Ruiz I (2000) Estimating genetic conformity between related ryegrass (*Lolium*) varieties. I. Morphology and biochemical characterisation. Mol Breed 6:569–580

Goodman MM, Bird RMcK (1977) The races of maize IV: tentative grouping of 219 Latin American races. Econ Bot 31:204–221

Gouesnard B, Dallard J, Bertin P, Boyat A, Charcosset A (2005) European maize landraces: genetic diversity, core collection definition and methodology of use. Maydica 50:225–234

Hammer K, Diederichsen A, Spahillari M (1999) Basic studies toward strategies for conservation of plant genetic resources. In: Serwinski J, Feberová I (eds) Proceedings of technical meeting on the methodology of the FAO World Information and Early Warning System on Plant Genetic Resources, Prague

Hammer Ø, Harper DAT, Ryan PD (2001) PAST: Paleontological statistics software package for education and data analysis Palaeontologia Electronica, vol 4, issue 1, art 4

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

Kresovich S, Williams JGK, McFerson JR, Routman EJ, Schaal BA (1992) Characterization of genetic identities and relationships of *Brassica oleracea* L. via a random amplified polymorphic DNA assay. Theor Appl Genet 85:190–196

Krauss SL (2000) Accurate gene diversity estimates from amplified fragment length polymorphism (AFLP) markers. Mol Ecol 9:1241–1245

Le Clerc V, Cadot V, Canadas M, Lallemand D, Guèrin D, Boulineau F (2006) Indicators to assess temporal genetic diversity in the French Catalogue: no losses for maize and peas. Theor Appl Genet 113:1197–1209

Lübberstedt T, Melchinger AE, Dussle C, Vuylsteke M, Kuiper M (2000) Relationship among early European maize inbreds. IV. Genetic diversity revealed with AFLP markers and comparison with RFLP, RAPD, and pedigree data. Crop Sci 40:783–791

Métais I, Hamon B, Jalouzot R, Peltier D (2002) Structure and level of genetic diversity in various bean types evidenced with microsatellite markers isolated from a genomic enriched library. Theor Appl Genet 104:1346–1352

McDermott JM, McDonald BA (1993) Gene flow in plant pathosystems. Annu Rev Phytopathol 31:353–373

Motto M, Lorenzoni C, Gentinetta E, Maggiore T, Salamini F (1978) Expected genetic gain for protein related traits and al location of resources in a modified opaque-2 synthetic. Maydica 23:35–44

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

Pollack LM (2003) The history and success of the public-private project on germplasm enhancement of maize (GEM). Adv Agron 78:45–87

Rebourg C, Gouesnard B, Welcker C, Dubreuil P, Chastanet M, Charcosset A (2003) Maize introduction into Europe: the history reviewed in the light of molecular data. Theor Appl Genet 106:895–903

Reif JC, Hamrit S, Heckenberger M, Shipprack W, Maurer HP, Bohn M, Melchinger AE (2005) Genetic structure and diversity of European flint maize populations determined with SSR analysis of individuals and bulks. Theor Appl Genet 111:906–913

Revilla P, Soengas P, Malvar RA, Cartea ME, Ordas A (1998) Isozyme variation and historical relationships among the maize races of Spain. Maydica 43:175–182

Rohlf FJ (1993) NTSYS-PC numerical taxonomy and multivariate analysis system. version 1.8. State Univ, New York

Singh SP, Gutiérrez JJA, Molina A, Urrea C, Gepts P (1991) Genetic diversity in cultivated common bean. II. Marker-based analysis of morphological and agronomic traits. Crop Sci 31:23–29

Smith JSC, Chin ECL, Shu H, Smith OS, Wall SJ, Senior ML, Mitchell SE, Kresovich S, Ziegle J (1997) An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.) comparisons with data from RFLPs and pedigree. Theor Appl Genet 95:163–173

Tommasini L, Batley J, Arnold GM, Cooke RJ, Donini P, Lee D, Law JR, Lowe C, Moule C, Trick M, Edwards KJ (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, Frijters 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