E

BMT-TWA/Maize/1/2
ORIGINAL: English
DATE: February 18, 2001

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

GENEVA

AD HOC CROP SUBGROUP ON MOLECULAR TECHNIQUES FOR MAIZE

First Session
Cambridge, United Kingdom, February 26 to 28, 2001

POTENTIAL USE OF MOLECULAR MARKERS FOR DISTINGUISHING MAIZE INBRED LINES AND HYBRIDS IN THE UPOV FRAME. A DRAFT PROPOSAL

Document prepared by experts from Spain

POTENTIAL USE OF MOLECULAR MARKERS FOR DISTINGUISHING MAIZE INBRED LINES AND HYBRIDS IN THE UPOV FRAME. A DRAFT PROPOSAL

J. Moreno-González
Centro de Investigaciones Agrarias de Mabegondo, Apartado 10, 15080 A Coruña, Spain

INTRODUCTION

Morphological descriptors of several phenotypic traits from the maize plant, including the kernel, ear, cob, tassel, stalk, silks, husks, leaves, inter-nodes, secondary roots, and anthocianic pigments are currently used for differentiating cultivars (mainly hybrids) to be included in the National List of commercial maize varieties. Similar descriptors are used to protect the breeder's right concerning the identity of parental maize inbred lines. Molecular markers, mainly isoenzimes, may be used as possible tools to evaluate deviations found in the variety identifications, as defined in the UPOV rules.

A large body of scientific data have been accumulated during the last fifteen years, which show the utility of biochemical and molecular markers (isoenzimes, RAPDS, RFLPs, SSRs, and AFLPs) for the purposes of identifying, fingerprinting and mapping the genotypes of different maize varieties (Smith 1988; Smith et al. 1990; Smith and Smith, 1991; Smith et al. 1990; Burstin et al. 1994; Hahn et al. 1995; Senior et al. 1996; Dhillmann et al. 1997, Dhillmann and Guérin, 1998; Smith et al. 1997; Senior et al. 1998; Pejic et al. 1998; Dubreuill and Charcosset 1999). Molecular markers have also confirmed to be very useful for helping in the selection and development of new cultivars (Stuber 1994; Ribaut and Hoisington 1998; Kraja and Dudley 2000). Several types of markers have been also used for the characterisation of maize inbreds, landraces and populations for genetic resources (Stuber and Moll 1983; Doebley et al. 1983; Khaler et al. 1986; Llauradó et al. 1993; Dhillmann and Guérin, 1998; Smith et al. 1997; Senior et al. 1998; Pejic et al. 1998).

The purpose of this report is to review published literature: (1) to know the state of art of molecular markers in relation to the identification of genotypes in maize; (2) to know the discriminating power of different types markers to distinguish among maize cultivars; and (3) to evaluate how useful are the different types of markers for being used in the distinctiveness of maize inbreds and commercial cultivars.

ATTRIBUTES OF MOLECULAR MARKERS

Molecular markers are fragments of DNA sequences from the plant nuclear chromosomes or the plant organelle genomes. Most of the utilised markers are from the chromosomes. They are located on some part of the genome and may be or may not be related to a specific gene, gene intron or expressed DNA sequence. Different kinds of molecular markers have been developed in the last years depending on the types of techniques and laboratory protocols used for their assessment. The length of the DNA sequence and its location, either at a random or at a specific and known tag site, are also factors that influence the type of markers.

Markers need to have some special qualities to be adopted for the identification of maize genotypes as routine laboratory techniques.

Some of these attributes are:

1. Neutral ADN fragments.
2. Codominance
3. Monolocus markers
4. Mendelian inheritance
5. High to moderate level of polymorphisms
6. Repeatability
7. High resolution
8. Correlation with phenotypic traits
9. Simple laboratory analytical techniques
10. High Efficiency results/costs

Some of these attributes are necessary to guaranty a high level of discrimination among cultivars, and to assure the stability and permanence of the procedure along the years. These characteristics may not be all reunited in one type of marker, however discussion should be focus on the best option available, although it may not be the optimum marker.

Attribute 1 will assure the absence of pleiotropic effects on phenotypes.

Attribute 2 will allow for fingerprinting heterozygous cultivars, such as maize hybrids, and to assess whether a particular genotype is a hybrid, an inbred line or a segregating population. This quality will also help to certify the identity of the inbred parents of a hybrid.

Attribute 3 will assign the DNA fragment to be located on a unique site of the maize genome and will provide a simple interpretation of the marker genetic analysis. This may not be easy to find in maize, since recent investigations suggest maize is a segmental allotetraploid, where some chromosomal regions may have a duplicate in another chromosome arm (Gaut and Doebly 1997).

Attribute 4 will ensure the marker inheritance behaves as a true gene and will provide a reasonable stability of markers in the genome through generations.

Attribute 5. A certain level of polymorphism is recommendable to easily find differences among cultivars, and to keep the number of markers at a manageable level. However a very high level of polymorphism (> 15 alleles per marker) may not be useful, because it will make more difficult to distinguish separation among bands, unless a specific software is developed for gel reading.

Attribute 6. Repeatability of band patterns for the same genotype is necessary among different laboratories. A desirable objective for a good marker should be to repeat identical profiles for the same maize inbred line or hybrid through generations.

Attribute 7. A neat and clear separation of bands is necessary for a quick and unequivocal reading of data. This will avoid misinterpretation when comparing similar profiles.

Attribute 8. An adequate correlation between molecular markers and the current UPOV morphological traits for differentiating varieties would be useful for helping to convert present UPOV system to a molecular-based system.

Attributes 9 and 10 would be desirable to facilitate the widespread of the method through UPOV members and to help adoption of molecular markers as routine and common techniques for cultivar identification in standard laboratories.

TYPES OF MARKERS

Different types of biochemical markers, such as isoenzimes, have been developed and used for studying maize variability since long time ago (Brown and Allard 1969). DNA molecular markers were developed later. Applications of these markers for genetic studies of maize have been so much diverse. Main uses include (1) assessment of genetic variability of populations; (2) characterisation of germplasm for evaluating the genetic resources; (3) identification and fingerprinting of genotypes; (4) estimation of genetic distances between populations, inbreds and breeding material; (5) aids for selection of qualitative traits, particularly through backcrossing; (6) detection of quantitative trait loci (QTL); (7) marker-assisted selection; and (8) identification of sequences of useful candidate genes, etc.

Brief descriptions of some of these markers are the following:

Isoenzymes. They were the first biohemical markers used to measure variability in plants. Studies of isoenzyme variation in maize populations for breeding purposes were reported long time ago (Brown and Allard 1969; Stuber and Moll 1972). Changes in isoenzyme gene frequencies were not clearly associated with the complex trait of yield (Stuber and Moll 1972). Isoenzymes have also been extensively used for characterising and classifying maize populations (Goodman and Stuber 1983; Doebley et al. 1983; Khaler et al. 1986; Llauradó et al. 1993). Advantages of this technique are the simplicity for being installed in any laboratory and the relatively cheap price of the analysis. One of the major disadvantages of isoenzymes for genotype identification and differentiating varieties is that the coverage of the maize genome is limited. Approximately, only 40 loci representing 21 isoenzymes may be routinely assayed (Lee 1994). In addition many enzyme loci have few allelic variants among inbreds, and some alleles have been found at extremely low frequencies, which may be not sufficient for assessing small differences among maize inbred lines.

Restricted fragment length polymorphisms (RFLPs). RFLPs are the different sizes of genomic DNA fragments caused by restriction enzymes and detected by specific hybridisation of homologous DNA probes. Technique for assessment of RFLPs is not simple enough to be installed in unspecialised routine laboratories. Protocols for this technique can be found in several publications (Gardiner, 1998; Karp et al. 1997). Basically, the technique consists in several phases: (1) extraction of DNA from maize young leaves (the following steps are included in this phase: sample lyophilisation, grinding, DNA isolation, and UV quantification of DNA); (2) restriction digest of genomic DNA by restriction endonuclease enzymes, such as *EcoR1*, *HindIII*, *BamH1*; (3) separation of DNA fragments by neutral agarose gel electrophoresis; (4) Southern blotting, i.e., transfer of gel DNA to a nylon or nitrocellulose filter; (5) baking the filter in an oven to immobilise the DNA; (6) hybridisation of filter to a labelled probe (the probe is cloned using radiolabelled or chemioluninescent nucleotides); (7) washing remains of non-hybridised probes; (8) development of filters to visualise fragment pattern of first probe; (9) washing and removal of denatured first probe to re-use the filter with a second probe.

A probe is a short DNA fragment (300 to 1400 bp) from a cloned expressed sequence, an unknown DNA fragment, or a part of a gene sequence of the maize genome. Restriction DNA fragments has larger size (2 to 25 kb) and depends on the restriction endonuclease enzyme used.

Advantages of RFLP markers are: (1) highly reproducible in different laboratories for the same genotype and the same enzyme-probe combination; (2) codominance behaviour; (3) Mendelian inheritance.

Disadvantages of RFLPs are: (1) large amount of extracted DNA is needed to perform analysis, which in some cases may preclude its use when an adequate amount of the plant can not be sampled; (2) time consuming and costs are high; (3) process is difficult to be automated into an integrated equipment, and complex to be performed in unspecialised laboratories.

RFLPs have been used in several studies to distinguish genotypes among a collection of maize inbred lines (Smith et al. 1990; Smith and Smith, 1991; Smith et al. 1990; Burstin et al. 1994; Hahn et al. 1995; 1996; Dhillmann et al. 1997, Dhillmann and Guérin, 1998; Pejic et al. 1998; Bernardo et al. 2000). In general, the RFLP method has shown to be very useful to separate different genotypes. Correlation between RFLPs and pedigree data were generally high (Table 1).

Random amplified polymorphic DNA (RAPD). RAPDs are very simple markers that can be detected by polymerase chain reaction (PCR) techniques. The RAPDs use a single arbitrary primer to start the PCR, resulting in the amplification of several DNA products from different locations in the genome. Each product will be derived from a region of the genome which share sequences similar to the primer and both ends are located on opposite strands sufficiently close for the amplification to work (Karp et al. 1997). Polymorphisms are detected as the presence or absence of bands, and they are determined by different sequence sizes caused from amplified segments between the end sites where the primer was bound. Products are separated on agarose gels in the presence of ethidium bromide and visualised under ultraviolet light

Advantages of this method are: (1) RAPD is very simple, quick and efficient method; (2) it only requires small equipment, such as a thermocycler and an electrophoresis apparatus.

Disadvantages of RAPDs are: (1) markers do not show co-dominance, because a potential band sequence will be always amplified either on homocigosis or heterocigosis; (2) RAPD profiles are difficult to reproduce among laboratories, even within the same laboratory when conditions or operators change, due in part to frequent miss-annealing; (3) each primer generally amplifies multilocus sequences.

RAPD markers have been used in several instances for the identification of maize inbreds (Kawata et al. 1995; Hahn et al. 1995; Pejic et al.1998). Genetic similarity (GS) based on RAPD markers was less correlated with co-ancestry coefficient (f_c) than the GS based on RFLP, AFLP or SSR markers (Messmer et al. 1993, Hahn et al. 1995; Pejic et al. 1998), (Table 1).

Amplified restriction fragment length polymorphism (AFLP). It is a technique that combines both, the restriction of genomic DNA as in RFLP and the PCR method as in RAPD. The procedure has been described by Vos et al. (1995). The first step involves digestion of

DNA with two restriction enzymes, one is a rare cutter and the another is a frequent cutter. The pair of restriction enzymes generally used are *Ecor I* and *Mse I*. The second step is the PCR amplification technique with a pair of primers designed to include nucleotide adapters to the restriction sites plus one additional nucleotide. The third step is a second PCR amplification on the previous PCR products, which is carried out with primers that include the same sequences than those used for the first PCR plus two additional nucleotides. Radio-labelled ATPs or chemioluminiscent molecules are linked to the Ecor I primer for visualising the products. The fourth step is the separation of amplified products on a polyacrylamide electrophoresis gel and visualisation with radio-labelled or staining procedures.

Advantages of AFLP are: (1) High polymorphism can be found in a single assay or a selected primer pair. (2) AFLPs are highly reproducible, as much as the RFLP technique. (3) Commercial kits and automation of the whole process are available for maize.

Disadvantages of AFLP are: (1) They do not show co-dominanace, however using gel scanners heterozygous may be identified; (2) the procedure is covered by patents for uses other than research; a specific licensed agreement is required for providing services to third parties; (3) it is a multilocus marker.

Limited data of AFLP markers are available for the identification of maize genotypes. However, AFLPs were high informative considering the low number of assays carried out (Pejic 1998, Lübberstedt et al. 2000).

Single sequence repeat (SSR) or microsatellites. SSRs are the latest type of markers so far developed, which have been used for identification of genotypes. They are amplified sequences of DNA regions with repeat nucleotide motifs located between two target primers. Basically, the procedure consists in a PCR that use target primer pairs, instead of a single random primer. A primer starts polymerisation in a target location of one strand from the 5' end and the other paired primer starts oppositely on other target location of the other strand from the 3' end, both ends separated about 80-300 bp apart. A large number of primer pairs have been developed for maize from different part of the genome. Primer length is about 15-30 bp. Presently, sequences of more than 1500 public primer pairs can be found in the Maize Data Base at the University of Missouri, USA. Some of these primer pairs are very informative and can detect a high number of alleles in a set of inbred lines.

Advantages of SSR markers are: (1) the method is relatively simple and can be automated, because it consists in a PCR followed by an electrophoresis gel analysis. (2) Most of the markes are monolocus and show Mendelian inheritance (Senior et al.1996, Smith et al.1997), however some markers were found to be duplicated in other regions of the genome, what is in agreement with the allotetraploid origin theory of maize (Gaut and Doebly 1997; Helentjaris 1995). (3) SSR markers are high informative, as determined by the high average polymorphism informative content (IPC) index (about 0.6) and the high average number of alleles per locus (about 5), after some primer pairs (≈ 40 %) were discarded due to low resolution (Smith et al. 1997; Senior et al. 1998). The IPC for a locus is estimated as $1- \Sigma p_k^2$, where p_k is the frequency of allele k at that locus. (4) Data of many public SSR markers can be retrieved from the maize Data Base (DB) for several characteristics, such as genome bin location, primer pair sequence, and gel profile images of the SSR marker for 10 US well-known maize inbred lines and one or two hybrids. In some instances, data are also available for the relationship and closeness of the particular SSR to other maize loci.

Disadvantages of SSR markers are: (1) SSRs are not generally linked to phenotypic traits. (2) A previous screening should be carried out to select adequate primer pairs with high resolution and high IPC. Data from the maize DB and a preliminary assay will help to make this selection. (3) No information is available about whether a particular SSR marker is able to yield stable profiles for different seed stocks of a stable phenotype inbred. This concern is pertinent because SSR polymorphisms can be detected by size differences as small as one base pair, when using acrylamide gels (Senior et al. 1998). Natural variation through generations may cause to make small changes in the SSR sequence of an inbred line. Then, probability of detecting a natural SSR change in a stable phenotype line is higher as the resolution power of the procedure is higher and the size of amplified SSR sequence is larger. A moderate resolution power, of about 2-3 % (in bp) of the SSR sequence size, may be more convenient to avoid excess of alleles per locus and to reduce the likely natural small changes of inbred gel profiles through generations, for the case where the inbred will remain phenotypically stable.

The SSR markers have shown to be as good or better than the other markers available so far, for the identification of maize genotypes (Smith et al.1997; Senior et al. 1998; and Pejic et al. 1998, Bernardo et al. 2000).

Other molecular markers. Very active research is presently going on the development of new markers to avoid shortcomings of the current DNA markers in maize. Specially, marker research is looking for targeting specific sites for expressed biochemical and physiological genes, phenotypic trait genes, QTLs, and specific candidate genes. New coming technology of markers will likely include sequence tag site (STS), expressed sequence tags (EST), and candidate genes.

PARAMETERS FOR ESTIMATING MOLECULAR-BASED GENETIC SIMILARITY

Several parameters have been proposed for estimating the genetic similarity between genotypes (S_{ij}).

Simple matching coefficient (Sneath and Sokal 1973):

$S_{ij} = (a + d)/(a + b + c + d)$, where a = number of "1-1" matches; b = number of "1-0" matches; c = number of "0-1" matches; and d = number of "0-0" matches; "1" means presence of an allele; and "0" means absence of this allele. This coefficient is useful when the whole bunch of alleles detected in a set of inbred lines are considered to establish similarity between two given inbreds, even when some of the alleles will not be directly involved in these particular lines.

Nei and Li's coefficient (Nei and Li 1979):

$S_{ij} = m/(m + n)$, where m = number of bands shared by inbreds i and j; and n = number of non-shared bands.

The modified Rogers' distance:

$$RD_{ij} = \frac{\left[\sum_{m=1}^{n}\sum_{k=1}^{l_m}(p_{imk} - p_{jmk})^2\right]}{2n}$$

where p_{imk} and p_{jmk} are the frequencies of allele k at locus m in inbreds i and j, respectively; l_m is the number of alleles at locus m; n is the number of loci. This coefficient is appropriate for populations and considers 0-0 matches. If used for inbred lines, p_{imk} and p_{jmk} will take values 0 or 1.

The coefficient of similarity using the Roger distance coefficient is estimated as 1 – *RD*.

The single matching coefficient (Sneath and Sokal 1973) and the Nei-Li (1979) coefficient are the most widely parameters used to compare the genetic similarity between two inbred lines depending on considering or not considering 0-0 matches, respectively.

RELATIONSHIP BETWEN PEDIGREE AND MOLECULAR DATA

Many works have been published for studying the relationship between pedigree similarity, as determined by the coefficient of coancestry (f_c), and the molecular-based genetic similarity (GS), as determined by the Nei-Li coefficient (1979), or the (1- *RD*) coefficient, in several types of maize material. The correlation coefficient (r_c) between the pedigree data (f_c) and the genetic similarity (GS) was low for RAPD markers (r_c = 0.40-0.49) and very high for RFLP markers (r_c = 0.57-0.95) for several type of material including hybrids, and US, European, flint and dent inbred lines (Table 1). Limited data exist showing the correlation coefficient (r_c) between the f_c based on pedigree and the GS based on SSR markers, however this r_c coefficient was similar to that between f_c and GS based on RFLPs for the available data (Tale 1).

Correlation coefficients between the GS based on RFLP and that based on other markers, such as RAPD and SSR are shown in Table 2. This coefficient between RFLP and SSR markers was high.

Bernardo et al. (2000) discussed the old concept of common bands (loci) that are *identical by descend* and common bands that are *alike in state but not identical by descend* to estimate the coefficient of coancestry based on molecular markers. They estimated the contribution of each parent to the descendants, using the simple matching coefficient (Sneath and Sokal 1973), when considering both the matching and mismatching bands among the involved individuals. Consider a cross of inbred lines *a* x *b*, and an inbred line *i* derived from that cross, the following relations hold (Bernardo et al. 2000):

$S_{ai} = \lambda_a + \lambda_b S_{ab}$
$S_{bi} = \lambda_b + \lambda_a S_{ab}$
$\lambda_a = (S_{ai} - S_{bi}S_{ab})/[1 - (S_{ab})^2]$
$\lambda_b = (S_{bi} - S_{ai}S_{ab})/[1 - (S_{ab})^2]$

Where, λ_a and λ_b are the genetic contribution of *a* and *b* to *i*, respectively; S_{ai}, S_{ab}, and S_{ab} are the marker similarities of *a* and *i*, *a* and *b*, and *b* and *i*, respectively.

In this way, Bernardo et al. (2000) were able to recalculate the coefficient of ancestry between two inbreds *i* and *j* (f_{ij}) as:

$$f_{ij} = \lambda_a f_{ai} + \lambda_b f_{bi}$$

The correlation coefficients (r_c) between pedigree and RFLP and SSR markers were 0.97 and 0.92, respectively; likewise the r_c between the RFLPs and the SSR was 0.87.

Evidence supports conclusions of several works that SSRs is likely the best option among present available markers for identification of cultivars (Smith et al. 1997, Senior et al. 1998, Pejic et al. 1998, Bernardo et al. 2000).

RELATIONSHIP BETWEEN MORPHOLOGICAL AND MOLECULAR DATA

Several works have studied the relationship between morphological traits and molecular markers (Stuber et al. 1992; Veldboom and Lee 1996a, and 1996b; Dillman et al. 1997; Kraja and Dudley 2000). The association between molecular markers and QTLs underlying morphological quantitative traits has allowed to assigning the location of some QTLs to specific regions of the maize genome (Beavis et al. 1994; Veldboom and Lee, 1996a and 1996b; Kraja and Dudley 2000). Furthermore, precise map location of many maize mutants is already known (Neuffer et al. 1997; maize DB, http://www.agron.missouri.edu).

Difficulties arise when trying to convert morphological data of a given maize phenotype, which is distinguished by UPOV descriptors, into molecular data. For example, the descriptor of heat units to pollen shedding behaves as a quantitative trait. Several QTLs for this trait, named *qhupol*, were detected on at least seven maize chromosomes when studying the F2:3 population derived from the Mo17 3 H99 cross (Veldboom and Lee 1996a). Some other different QTLs for the same trait were detected in a population derived from the Mo17 3 B73 cross (Beavis et al. 1994). The maize DB enlists 32 QTLs *qhupol*1-32. Also, the magnitude of the additive and dominance effect of these QTLs depended on the environment where genotypes were tested. In addition, no sequences or probes for these genes have been published so far. Thus, the description of the trait of heat units to pollen by molecular data is not possible yet.

Other traits, such as kernel row and plant height, are also controlled by several QTLs, which can not been described by molecular markers, because no direct probes for all these QTLs have been developed so far. In certain cases, QTLs have been located in the same chromosomal regions than loci defined by alleles with qualitative effects (mutants) for the same trait. For example, a QTL for plant height has been found associated to the RFLP probe *an1*, which in turn is at or tightly linked to the *anther ear* locus responsible for the phenotype mutant of dwarf or intermediate stature of the plant (Veldboom and Lee 1996b).

The work of Dillman et al. (1997) showed that the relationship between molecular and morphological distance was not clear when 145 inbred lines were evaluated using both RFLP markers and a set of discriminant morphological traits.

In contrast, some qualitative traits, such as the endosperm types *waxy* (*wx*), *sugary1* (*su1*), *opaque2* (*o2*) and many others, have been cloned and there are either SSR or RFLP probes for direct identification. For example, the *o2* gene has been cloned (Schmidt et al. 1987) and a

SSR marker *phi057* (bin 7.01) has been detected for this gene (Chin et al. 1996). Likewise, SSR markers, such as *phi027* (bin 9.03) and *phi022*, are tightly linked to the *waxy1* (*wax1*) locus.

Therefore, it can be concluded that molecular characterisation of the UPOV morphological traits showing quantitative variation, used for distinguishing maize cultivars, will be a very complex or impossible task with the tools presently available. By the contrary, some of the UPOV qualitative traits may be characterised or detected by molecular markers, such as SSR or RFLP probes.

CHOICE OF THE TYPE OF MOLECULAR MARKER FOR DIFFERENTIATING CULTIVARS

Choice of the type of molecular markers.
According to the reviewed data, it is suggested to choose SSR markers as the first option for the identification and differentiation of maize inbreds and hybrids. Reasons for this choice are: (1). The SSR assay is a simple procedure. Ninety-six genotypes can be simultaneously assayed for each primer pair through the entire process, including the PCR reactions and the electrophoresis analysis, with technology already available (Senior et al. 1998). Furthermore, multiplex amplification technology may allow for simultaneous PCR of more than one SSR locus (Mitchell et al. 1997). (2) The averaged polymorphism for SSR markers was high, about 5 alleles per locus (Senior et al. 1998; Smith et al. 1997). (3) Generally, the SSR assay is a monolocus system with Mendelian inheritance behaviour. The frequency of more than one locus per primer pair was low (0.02) and smaller than for RFLP markers (0.04) (Senior et al. 1998; Smith et al. 1997). (4) The SSRs show co-dominance, which is essential to identify a hybrid and its parental lines. (5) A high number of public SSR primer pairs are available (more than 2000 so far). Also, SSR markers can be precisely located on the maize genome (maize DB, *http://www.agron.missouri.edu*). (6) Some of the available SSR primers are close to loci controlling qualitative traits or mutants. (7) Effective cost per genotype and primer is similar to that for RAPD markers and much smaller than for RFLPs.

Second choice of markers for fingerprinting and identification of genotypes will be the RFLPs, because the above reasons 2, 3, 4, 5, and 6 for SSRs may also be applicable for the RFLPs.

A PROPOSAL FOR SELECTING A GROUP OF PERMANENT SSR MARKERS

A group of only five highly polymorphic SSR markers provided unique genotype profiles for differentiating 94 maize inbreds (Senior at al. 1998). However, it is thought that a dense coverage of the genome will be necessary to guarantee distinctiveness among the wide universe of present and future inbred lines to be protected.

A scheme to select a set of markers for distinguishing maize inbreds is proposed for discussion among the UPOV members:

Step 1. Identification and map location of genes controlling the qualitative phenotype traits currently used for differentiating maize inbred lines to be registered in the National Lists for protection purposes. Some of the traits are:

(a) Anthocianic pigments in the cotiledons, anther, glumes, cob, pericarp, silks, leaves, and stalk. Some of the genes controlling these traits include the *A1*, *A2*, *B1*, *C2*, *R*, *Pl1*, *Bz* series (Coe 1994, Neuffer et al. 1997).
(b) Yellow (*Y*) and white (*y*) endosperm colour.

Step 2. Identification and map location of genes controlling qualitative phenotype traits with agronomic, nutritive or industrial interest. Some of these genes may include:
Opaque2 (*o2*), *waxy* (*wx1*), *sugary1* (*su1*), *shrunken2* (*sh2*), *brachyitic2* (*br2*), *brownmidrib* (*br1*), *liguless1* (*lg1*), *Dwarf8-1* (*D8-1*), (Neuffer et al. 1997).

Step 3. Identification and map location of important and consolidated QTLs responsible for controlling quantitative traits used for differentiating inbred lines to be registered in the National Lists. Some of these traits may be plant height, ear height, leaf angle, ear row number, ear diameter, kernel type, leaf number, leaf length, heat units to pollen, etc. (Part of this information can be retrieved from the maize DB at the University of Missouri, *http://www.agron.missouri.edu/*).

Step 4. Searching for SSR primer pairs located on map positions close to the identified and selected maize genes and QTLs (maize DB, *http://www.agron.missouri.edu;* Neuffer et al. 1997).

Step 5. Complement the set of markers with additional SSR primer pairs in order to have a dense and even coverage of the maize genome (maize DB, *http://www.agron.missouri.edu/*). This set may include SSR markers already tested to have a high IPC and a good resolution in previous works (Smith et al. 1997, Senior et al. 1998).

Step 6. An exploratory study with an initial set of 150-200 SSR markers should be carried out for assessing 400-500 genotypes, including both inbred lines and hybrids, from wide origin. Several isogenic line pairs, each differentiating in one or two qualitative mutants, such as *o2*, *br2*, *fl2*, *su1*, *lg1*, *wx1*, *bm1*, *y*, etc, can be included. Pairs of isogenic inbred lines may also differentiate in one or more disease resistance genes, such as *Ht*. Normal inbreds and their counterpart inbreds converted to specific characteristics, such as earliness, lodging resistance, shorter plants, etc., may be also included. Pedigree related inbreds with coefficient of coancestry about 0.75, 0.87, 0.93 and 0.97, corresponding to 1, 2, 3 and 4 backcrosses respectively, should be also incorporated into the study.

A group of hybrid crosses among some of the inbred lines should be added to the analysis to test the Mendelian inheritance of the SSR markers.
The complete study or a part of that should be repeated in two laboratories to check the repeatability of the SSR markers.

Step 7. Presence or absence of a band for a given genotype at each specific allele of each SSR locus will be denoted by “1” or “0”, respectively. A matrix with the 1 and 0 elements will be created, where rows correspond to genotypes and columns correspond to all alleles of all tested SSR loci. Genetic similarity between any two genotypes may be estimated using the simple matching coefficient (Sneath and Sokal 1973).

Step 8. Select a subset of permanent SSR markers, that will efficiently discriminate the inbred lines, for future genotype assessment. The criteria for choosing the subset of SSRs may include (1) elimination of markers showing non-polymorphism, non-Mendelian inheritance, and multilocus behaviour. (2) Choice of a first group of markers showing high IPCs. (3)

Development of an algorithm for aggregating SSR markers to this group, in such way that the genetic distance for any pair of given inbreds, and the correlation between genetic similarity and coefficient of coancestry will be maximised. (4) Complementation of the group with other markers in order to have an adequate coverage of the genome.

MAXIMUM GENETIC SIMILARITY TO DECLARE DIFFERENT TWO INBRED LINES

When comparing molecular bands of two inbred lines to estimate genetic similarity, some of the common bands may be *identical by descend*, while the others may be *alike in state but not identical by descend.* The problem will be to know which proportion of the common bands are *identical by descend* in relation to the bands *alike in state*. Thus, it is expected that genetic similarity (GS) based on observed molecular bands will be higher than the GS based on pedigree coefficient of coancestry (f_c), because the GS of molecular markers will likely include both kinds of bands, while the pedigree GS presumably only includes alleles identical by descend. However, data from the work of Smith et al. (1997) show that the genetic distance based on pedigree, determined as $1 - f_c$, was similar to that based on RFLP and SSR markers using the Nei-Li coefficient (1979) for the high-related inbred lines. Nevertheless, the pedigree genetic distance was a little higher than that based on molecular markers for the low-related lines, as expected.

In addition, the work of Bernardo et al. (2000) showed that the correlation coefficient (r_c) between the coefficient of coancestry (f_c) based on pedigree and the f_c based on SSR markers was 0.92. Although some pairs of inbreds showed significant differences between the f_c based on pedigree and SSR, the average difference over all pairs of inbreds was small, 0.01, i.e., 0.42 and 0.41 for the average f_c of SSR and pedigree, respectively.

Therefore, it seems that GS based on SSR markers can be considered as an adequate parameter to study the relationships among inbreds.

Thus, a given inbred line to be declared different from each one already registered in the Maize Inbred Register Molecular Data Base (MIRMDB) has to meet one of the following requirements:

1. The genetic similarity (GS) between the candidate inbred and any other in the permanent MIRDMB, as determined by the Nei-Li coefficient (1979) based on molecular markers, has to be lower than a fixed level (GSmax). The fixed level may be 0.87 + SE or 0.93 + SE or 0.96 + SE, corresponding to two, three or four backcrosses plus one or two standard errors (SE), respectively. The standard error may be obtained either as $[GS(1 - GS)]^{0.5}/ n$, where n is the number of markers or by the *bootstrapping statistical technique*.

2. If the GS of the candidate inbred is higher than the fixed GSmax, this inbred should proof to be different for a specific and valuable characteristic, such as a DNA sequence controlling a disease or pest resistance or a special trait. Markers or probes for assessing this characteristic may be provided to the Organisation in charge of the inbred identification.

IDENTIFICATION OF DIFFERENT HYBRIDS

Two different approaches seem to be important to consider for distinguishing the identity of maize hybrids.

1. Prior molecular identification and registration of the involved parental inbred lines in the MIRDMB, showing that the combination of the hybrid cross between the parental inbreds has not been previously registered in the Maize Hybrid Register Molecular Data Base (MHRMDB). Molecular identification of the hybrid genotype must be also performed to assess the identity of the parents.

2. Exclusive molecular identification of the hybrid cross, without considering prior molecular identity of the inbred parents in the MIRMDB. The GS between the given hybrid and any other already registered in the MHRMDB can be estimated using the same type of molecular markers and the same Nei-Li coefficient than those described for the inbred lines. A maximum threshold level for the GS need also to be fixed for declaring the hybrid different.

Table 1. correlation coefficients (*r*) between coefficient of coencestry (f_c) based on pedigree data and genetic similarity (gs) based on molecular markers, RAPD, AFLP, RFLP, SSR.

Reference	Type of cultivar	RAPD	AFLP	RFLP	SSR
Smith et al. 1990	Hybrid			0.90	
Smith and Smith 1991	Hybrid			0.91	
Smith and Smith 1992	Hybrid			0.95	
Messmer et al. 1993	Flint inbred			0.84	
Messmer et al. 1993	Dent inbred			0.91	
Burstin et al. 1994	Inbred			0.63	
Hahn et al. 1995	Flint inbred	0.42			
Hahn et al. 1995	Dent inbred	0.49			
Pejic et al. 1998	Inbred	0.40	0.62	0.57	0.52
Smith et al. 1997	Inbred			0.80	0.81
Lübberstedt et al. 2000	Flint inbred		0.79		
Lübberstedt et al. 2000	Dent inbred		0.69		

Table 2. Correlation coefficients (*r*) between the genetic similarity (GS) based on RFLP markers and GS based on RAPD or SSR markers.

Reference	Type of cultivar	RAPD	SSR
Hahn et al. 1995	Flint inbred	0.47	
Hahn et al. 1995	Dent inbred	0.26	
Smith et al. 1997	Inbred		0.85
Pejic et al. 1998	Inbred	0.51	0.59

REFERENCES

Beavis, W.D., O.S. Smith, D. Grant, and R. Ficher. 1994. Identification of quantitative trait loci using a small sample of topcrosses and F4 progeny from maize. Crop Sci. 34:882-896.

Bernardo, R., J. Romero-Sverson, J. Ziegle, J. Hauser, L. Joe, G. Hookstra, and R.W. Doerge. 2000. Parental contribution and coefficient of coancestry among inbreds: pedigree, RFLP, and SSR data. Theor Appl Genet 100:552-556.

Brown, A.H.D., and R.W. Allard.1969. Inheritance of isozyme differences among the inbred parents of a reciprocal recurrent selection of maize. Crop Sci. 9: 72-75.

Burstin, J., D. De Vienne, P. Dubreuil, and C. Damerval.1994. Molecular markers and protein quantities as genetic descriptors in maize. I. Genetic diversity among 21 inbred lines. Theor. Appl. Genet. 89: 943-950.

Chin, E., L. Senior, H. Shu, and J.S.C. Smith. 1996.Maize simple repetitive DNA sequences: Abundance and allelic variation. Genome 39:866-873.

Coe, E.H. 1994. Anthocianin Genetics. *In* The maize handbook, M. Freeling and V. Walbot, eds., pp. 279-282, Springer-Verlag, New York, Inc.

Dillmann, C., A. Ver-Hen, D. Guerin, A. Charcosset, and A. Murgneux. 1997. Comparison of RFLP and morphological distance between maize (Zea mays, L.) inbred lines. Consequences for germplasm protection purposes.. Theor. Appl. Genet. 95: 92-102.

Dillmann, C., and D. Guérin. 1998. Comparison between maize inbred lines: Genetic distances in the expert's eye. Agronomie 18: 659-667.

Doebley, J.F., M.M. Goodman, and C.W. Stuber. 1983. Isozyme variation in maize from the Southwestern United States: Taxonomic and anthropological implications. Maydica 28: 97-120.

Dubreuill, P., and A. Charcosset. 1999. Relationships among maize inbred lines and populations from European and Nort-America origins as estimated using RFLP markers. Theor. Appl. Genet. 99: 473-480.

Gardiner, J. M., 1998. UMC Maize RFLP Procedures Manual. Unpublished. (A compilation of procedures with modifications, updated 1998).

Gaut, B.S. and J.F. Doebley. 1997. DNA sequence evidence for the segmental allotetraploid origin of maize. Proc. Nat. Acad. Sci. USA 94: 6809-6814.

Goodman, M.M. and C.W. Stuber. 1983. Races of maize. VI. Isozyme variation among races of maize in Bolivia. Maydica 28: 169-187.

Hahn, V., K. Blankernborn, M. Schwall, and A.E. Melchinger. 1995. Relationships among early European maize inbreds: III. Genetic diversity revealed with RAPD markers and Comparison with RFLP and pedigree data. Maydica 40: 299-310.

Helentjaris, T. 1995. Atlas of duplicated sequences. Maize Genet. Coop. Newsl. 69:67:81.

Karp, A., S. Kresovich, K.V. Bhat, W.G. Ayad, and T, Hodgkin. 1997. Molecular tools in plant genetic resources conservation: a guide to technologies. IPGRI Technical Bulletin No. 2. International Plant Genetic Resources Institute, Rome, Italy.

Kawata, M., S. Yazaki, Y. Monma, Y. Miura, F. Takaiwa, Y. Shimamoto, and S. Ueta. 1995. Pedigree assessment using random aplified polymorphic DNA in maize (*Zea mays* L. Inbred lines. Grassland Science 41:251-255.

Khaler, A.L., A.R. Hallauer, and C.D. Gardner. 1986. Allozyme polymorphisms within and among open-pollinated and adapted exotic populations of maize. Theor. Appl. Genet. 75: 592-601.

Kraja, A.T., and J.W. Dudley. 2000. QTL analysis of two maize inbred line crosses. Maydica 45:1-12.

Lee, M. 1994. Inbred lines of maize and their molecular markers. *In* The maize handbook, M. Freeling and V. Walbot, eds., Springer-Verlag, New York, Inc.

Llauradó, M. J. Moreno-Gonzalez, and P. Arús. 1993. Classification of Northern Spanish populations of maize by numerical taxonomy. II. Isozyme variation. MAYDICA 38: 249-258.

Lübberstedt T., A.E. Melchinger, C. DuBle, M. Vuylsteke, M. Kuiper. 2000. Relationships among early european maize inbreds:IV. Genetic diversity revealed with AFLP markers and comparison with RFLP, RAPD, and pedigree data. Crop Science 40: 783-791

Messmer, M.M, A.E. Melchinger, R.G. Herrmann, and J. Boppenmaier. 1993. Relationship among early European maize inbreds. II. Comparison of pedigree and RFLP data. Crop Sci. 33: 944-950.

Mitchell, S.E., S. Kresovich, C.A. Jester, C.J. Hernandez, and A.K. Szewc-McFaden. 1997. Amplification of multiplex PCR and fluorescence-based, semi-automated allele sizing technology for genotyping genetic resources. Crop Sci. 37:617-624.

Nei, M, and W. Li. 1979. Mathematicalmodel for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA 76:5269-5273.

Neuffer, M.G., E.H. Coe, and S. Wessler 1997. The mutants of maize. Cold Spring Harbor Press.

Pejic, I., P. Ajmone-Marsan, M. Morgante, V. Kozumplik, P. Castiglioni, G. Taramino and M. Motto. 1998. Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs, and AFLPs. Theor. Appl. Genet. 97: 1248-1255.

Ribaut, J.M., and D. Hoisington. 1998. Marker-assisted selection: new tools and strategies. Trends in Plant Science 3: 236-239.

Senior, M.L.; E.C.L. Chin, M. Lee, J.S.C. Smith, and C.W. Stuber. 1996. Simple sequence repeat markers developed from maize sequences found in the GenBank database: map construction. Crop Sci. 36: 1676-1683.

Schmidt, R.J., F.A. Burr, and B. Burr. 1987. Transposon tagging and molecular analysis of the maize regulatory locus *opaque2*. Science 238:960-963.

Senior, M.L., J.P. Murphy, M. M. Goodman, and C.W. Stuber. 1998. Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. Crop Sci. 38: 1088-1098.

Smith, J.S.C. 1988. The diversity of U.S. hybrid maize germplasm; isozyimic and chromatografic evidence. Crop Sci. 28: 63-70.

Smith, O.S., J.S.C. Smith, S.L. Bowen, R.A. Tenborg, and S.J. Wall. 1990. Similarities among a group of elite maize inbreds as measured by pedigree, F_1 grain yield, grain yield, heterosis and RFLP's. Theor. Appl. Genet. 80: 833-840.

Smith, J.S.C., O.S. Smith. 1991. Restriction fragment length polymorphisms can differenciate among U.S. maize hybrids, Crop Sci. 31: 893-899.

Smith, J.S.C., O.S. Smith, S.L. Bowen, R.A. Tenborg, S.J. Wall. 1991. The description and assessment of distances between inbred lines of maize. III. A revised scheme for the testing of distinctiveness between inbred lines utilizing DNA RFLP's. Maydica 36: 213-226.

Smith, J.S.C., E.C.L. Chin, H. Shu, O.S. Smith, S.J. Wall, M.L. Senior, S.E. Mitchell, S. Kresovich, and J. Ziegle. 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.

Snath, P.H.A., and R.R. sokal. 1973. Numerical taxonomy. W.H. Freeman and Co. San Francisco, USA.

Stuber, C.W. 1995. Mapping and manipulating quantitative traits in maize. Trends in Genetics 11: 477-481.

Stuber, C.W. and R.H. Moll. 1972. Frequency changes of isozyme alleles in a selection experiment for grain yield in maize (*Zea mays* L.). Crop Sci. 12: 337-340.

Stuber, C.W. and M.M. Goodman. 1983. Allozyme genotypes for popular and historically important lines of corn. U.S. Agric. Res. Serv. Results, Southern, Ser., no.16.

Stuber, C.W., S.E. Lincoln, D.W. Wolf, T. Helentjaris, and E.S. lander. 1992. Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132: 823-829.

Veldboom, L.R., and M. Lee. 1996a. Genetic mapping of quantitative trait loci in maize in stress and nonstress environments: I. Yield and yield components. Crop Sci.36:1310-1319.

Veldboom, L.R., and M. Lee. 1996b. Genetic mapping of quantitative trait loci in maize in stress and nonstress environments: II Plant height and flowering. Crop Sci.36:1320-1327.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. Van De Lee, M. Hornes, A. Frijters, J. Pot, J. Poleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23:4407-4414.

[End of document]