Aus dem Institut
für Pflanzenzüchtung, Saatgutforschung und Populationsgenetik
der Universität Hohenheim
Fachgebiet: Angewandte Genetik und Pflanzenzüchtung
Prof. Dr. A. E. Melchinger

Identification of essentially derived varieties in maize (*Zea mays* L.) using molecular markers, morphological traits, and heterosis

Dissertation
zur Erlangung des Grades eines Doktors
der Agrarwissenschaften
der Fakultät Agrarwissenschaften
der Universität Hohenheim

von

Dipl.-Ing. sc. agr.
Martin Heckenberger
aus Ochsenhausen - Reinstetten

2004

Die vorliegende Arbeit wurde am 30. Januar 2004 von der Fakultät Agrarwissenschaften der Universität Hohenheim als „Dissertation zur Erlangung des Grades eines Doktors der Agrarwissenschaften (Dr. sc. agr.)“ angenommen.

Tag der mündlichen Prüfung:	19. Februar 2004
1. Prodekan:	Prof. Dr. K. Stahr
Berichterstatter, 1. Prüfer:	Prof. Dr. A.E. Melchinger
Mitberichterstatter, 2. Prüfer:	Prof. Dr. H.-P. Piepho
3. Prüfer:	Prof. Dr. R. Blaich

"… it cannot be now many years, if the investigations go on at the present rate, before the breeder will be in a position not so very different from that in which the chemist is:… when he will be able to do what he wants to do, instead of merely what happens to turn up."

William Bateson, "Practical Aspects of the New Discoveries in Heredity."
Address to the international Conference on Plant Breeding and Hybridization,
New York City, Aug. 30-Sept. 2, 1902

Contents

[1] M. Heckenberger, M. Bohn, J.S. Ziegle, L.K. Joe, J.D. Hauser, M. Hutton and A.E. Melchinger. 2002. Variation of DNA fingerprints among accessions within maize inbred lines and implications for identification of essentially derived varieties. I. Genetic and technical sources of variation in SSR data. Mol. Breed. 10:181-191.

[2] M. Heckenberger, J. Rouppe van der Voort, A.E. Melchinger, J. Peleman and M. Bohn. 2003. Variation of DNA fingerprints among accessions within maize inbred lines and implications for identification of essentially derived varieties: II. Genetic and technical sources of variation in AFLP data and comparison with SSR data. Mol. Breed. 12:97-106.

[3] M. Heckenberger, M. Bohn, and A. E. Melchinger. 2004a. Identification of essentially derived varieties (EDVs) derived from biparental crosses of homozygous lines. II. SSR data from maize inbreds Crop Sci. submitted.

[4] M. Heckenberger, M. Bohn, D. Klein, and A. E. Melchinger. 2004b. Identification of essentially derived varieties (EDVs) derived from biparental crosses of homozygous lines. III. Morphological distances and heterosis in comparison with SSR and AFLP data in maize Crop Sci. submitted.

Abbreviations

AFLP	amplified fragment length polymorphism
ais	alike in state
ASSINSEL	International Association of Plant Breeders
ASTA	American Seed Trade Association
BC	backcross
DNA	deoxyribonucleic acid
DUS	distinctness, uniformity, stability
EDV	essentially derived variety
f	coancestry coefficient
GD	genetic distance
ibd	identical by descent
IDV	independently derived variety
ISF	International Seed Federation
IV	initial variety
MD	morphological distance
MPH	mid-parent heterosis
NPA	non-parental allele
oais	only alike in state
PCR	polymerase chain reaction
PVP	plant varietal protection
RAPD	random amplified polymorphic DNA
RFLP	restriction fragment length polymorphism
SEPROMA	Chambre Syndicale des Entreprises Françaises de Semences de Maïs
SNP	single nucleotide polymorphism
SSR	simple sequence repeat
UPOV	Union for the protection of new varieties of plants

1. General Introduction

The Need for new Plant Varieties

Since the 1960s, the human world population has more than doubled from 3 billion to currently more than 6 billion and it is estimated to reach 8 billion by the year 2025 (Vaupel et al., 1998). At the same time, the amount of arable land is decreasing due to diversion to nonfarm uses, such as urban or industrial development or natural phenomena such as expanding deserts. Therefore, the only option is to produce more food on less land to meet the needs of the increasing population. Thus, the critical importance of ecologically sustainable advances in the productivity and profitability of major farming systems is evident. At present, the average yield increase of major crops such as maize (*Zea mays* L.) (Duvick and Cassman, 1999), rice (*Oryza sativa* L.) (Mann, 1997; Vasil, 1998), soybean (*Glycine max* L.) (Ustun et al., 2001) or wheat (*Triticum aestivum* L.) (Reynolds et al., 1999) amounts to approx. 0.5 - 1.5% per year. This yield increase is the result of improvements in soil cultivation, fertilization, and plant protection, but also to a large extent by continuous breeding progress made by plant breeders for all crops, because 50% of the yearly yield increase in crop production is attributed to genetic modification and selection. Therefore, plant breeding is playing a key role in this process, because the average yield increase of major crops is currently slightly lower than the growth rate of the world population of currently approx. 1.8% *p.a.* (Daily et al., 1998). Consequently, the particular importance of the continuous development of new plant varieties in all major crops is obvious to contribute to improvements in quantity, quality, and cost of satisfying the requirements for food, fuel, fiber and for raw materials for industry (Tanksley and McCouch, 1997).

The Need for legal Protection of new Plant Varieties

Plant breeding aims at modifying plants genetically in a way to adapt them better to the specified needs of the people (Becker, 1993) and requires extensive crossing programs and sophisticated selection strategies. Thus, procedures of breeding new crop varieties are time-consuming and expensive, *e.g.,* the development of a new maize hybrid takes up to 15

years and requires monetary investments of up to 5.000.000 US$ (Troyer et al., 2002). Expenditures of more than one billion US$ were achieved in 2000 only by US plant breeding companies, combined with a research and development quota (R&D-quota) of approx. 15%, which makes plant breeding one of the most research intensive business fields (Heisey et al., 1999).

Biological organisms including plants are characterized by the ability to reproduce or replicate itself. This attribute faces plant breeders with special problems since the release of propagating material of variety to growers enables them in many cases to reproduce the variety without further recourse to the breeder. In addition, the breeder's competitors can secure supplies of propagating material and compete in a short time profiting from the breeding effort of the breeder of the original variety over many years. In order to secure an appropriate reward for the investments made, it is important to protect plant varieties in the best possible way against plagiarism and misuse as intellectual properties of the breeders (International Seed Federation, 2002). This protection can either be warranted by requesting patent protection for new plant varieties, as is possible *e.g.* in the USA, or by applying for plant varietal protection (PVP).

Plant varietal Protection within the UPOV-System

Besides patents, many countries have established systems whereby exclusive rights of exploitation are granted to the breeders of new varieties in order to provide breeders with an opportunity to receive a reasonable return on past investments. Furthermore, PVP systems were implemented to provide an incentive for continued or increased investment in the future and to recognize the moral right of the innovator to be recognized as such and his economic right to receive remuneration for his efforts.

The International Union for the Protection of New Varieties of Plants (UPOV) is a union of states which have agreed to grant exclusive rights of exploitation to the breeders of new plant varieties on an internationally harmonized basis. Currently, 52 countries joined the UPOV convention (UPOV, 1961) and implemented their legal regulations into their national PVP acts. Accordingly, a variety needs to be distinguishable from all other varieties of the same crop, uniform, new, and stable to receive the privilege of PVP, as

assessed with the testing procedures for distinctness, uniformity and stability (DUS). These well-defined procedures for DUS-testing relying on morphological and biochemical traits are the basis of PVP for numerous crops.

The "Breeders' Exemption"

Apart from the protection against misuse, however, a variety needs to be freely available to all breeders as a germplasm resource to secure sustainable breeding progress. Therefore, the principle of the "breeder's exemption" was implemented into the 1978 act of the UPOV convention (UPOV, 1978), enabling breeders to use protected varieties for the purpose of producing new varieties. Accordingly, any protected variety may be freely used as a source of initial variation in breeding programs to develop further varieties and any such variety may itself be protected and exploited without any obligation on the part of its breeder.

For the purpose of breeding new varieties, the breeder's exemption is a fundamental aspect of the UPOV system of PVP and represents the main difference between PVP and patents. It warrants a continuous breeding progress on the basis of the PVP system and recognizes that real progress in breeding, which for the benefit of society must be the goal of intellectual property rights in this field, relies on access to the latest improvements and new variation. Therefore, access is warranted to all breeding materials in the form of modern varieties, as well as landraces and wild species, to achieve the greatest progress and is only possible if protected varieties are available for further breeding.

Misuse of the Breeders' Exemption

Recently, newly developed methods and strategies have created the technical basis for a misuse of the breeder's exemption. Among others, these methods comprise repeated (marker-assisted) backcrossing to protected varieties, the addition of one or few genes to protected varieties by genetic engineering, reselection within protected varieties (*e.g.,* selection for natural or induced mutants within ornamental crops) and the generation of "cosmetic" variation within a protected variety, (*e.g.*, generation of somaclonal variants by repeated tissue culture cycles or selection of clones within synthetics) (Knaak et al., 1996).

These techniques provide the possibility to undermine the PVP system in the original intention of the UPOV convention, which implied to support classical, creative plant breeding. Therefore, it has become possible to plagiarize protected varieties without indemnification for the breeders of the corresponding original varieties and very little genetic variation is created.

Essentially Derived Varieties

The concept of essentially derived varieties (EDVs) was implemented into the 1991 act of the UPOV convention to prevent plagiarized breeding and to allow at the same time the use of protected varieties for the development of new cultivars. A variety is deemed to be essentially derived from an initial variety (IV) when it is (i) predominantly derived from the IV, or from a variety that is itself predominantly derived from the IV, while retaining the expression of the essential characteristics that result from the genotype or combination of genotypes of the initial variety, (ii) it is clearly distinguishable from the initial variety and (iii) except for differences, which result from the act of derivation, it conforms to the initial variety in the expression of essential characteristics that result from the genotype or combination of genotypes of the initial variety (UPOV, 1991).

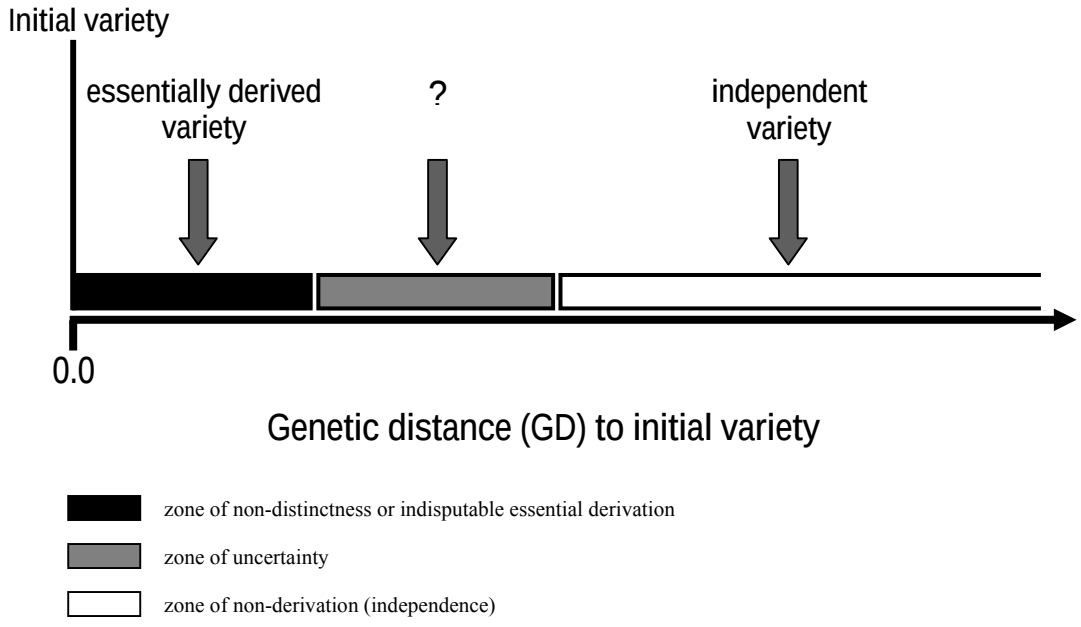


Figure 1: The concept of different thresholds to discriminate between essentially derived and independent varieties based on their GD according to an initial variety.

According to ASSINSEL, (1999a), a worldwide association of plant breeders, a variety is essentially derived if all three criteria are met. If one of the criteria is not fulfilled, there is no essential derivation. In contrast to DUS testing for PVP, which is performed by governmental organizations, the breeder of the IV is supposed to supply evidence that a new variety was essentially derived from the IV. To prove a predominant derivation, either a directly documented evidence will be necessary (Gilliland et al., 2000), or forensic approaches will have to be applied, similar to those applied in human genetics for detecting parentage (Gill et al., 1995; Graham et al., 2000). Distinctness will most likely be observed by DUS-testing, however, no consensus has currently been achieved regarding the methods to be used for determining the expression of the essential characteristics and to observe the genetic conformity between IVs and EDVs.

In addition, the breeders have not yet agreed on accepted or non-accepted breeding procedures yielding independently derived varieties (IDVs) or EDVs, respectively. In particular, the number of acceptable backcrosses to a protected variety without generating an EDV has not been fixed for all major crops including maize. Therefore, appropriate methods to identify EDVs have not been defined, because a detailed comparison of all eligible identification methods has not been performed so far. Theoretically, genetic distances (GDs) based on molecular markers, morphological traits as used for DUS testing, probabilistic values such as the coancestry coefficient (Malécot, 1948), or other descriptors like heterosis or combining ability could serve for identification of EDVs because they all provide information on the relatedness of cultivars (International Seed Federation, 2002). Due to this lack of knowledge, thresholds to distinguish between IVs and EDVs have not yet been defined (Fig. 1.). Thus, the EDV concept was implemented into practical plant breeding only for very few crops, such as lettuce (International Seed Federation, 2003), but is still not used for all major crops, including maize.

Identification of EDVs

As the concept of essential derivation is rather based on genotypes than on phenotypes, the use of scientifically reliable methods to observe GDs is required by breeding organizations, such as ASSINSEL, (1999b). Therefore, justiciable criteria have to be de-

veloped and validated because of the legal consequences for infringing PVP granted within the UPOV-system.

Genetic fingerprints based on molecular marker data proved to be reliable tools to determine parentage and to identify crime suspects in humans (Gill et al., 1995; Graham et al., 2000). Like in criminology, molecular markers became key instruments in the plant breeding sector to evaluate the illegal use of germplasm (Martin et al., 1991). Therefore, molecular markers were proposed repeatedly as suitable tools to identify essential derivation in all major crops (ASSINSEL, 1999a; Bernardo and Kahler, 2001; Knaak et al., 1996; Roldan Ruiz et al., 2000; Smith and Smith, 1989b), as they provide a direct measure of the true relatedness of two cultivars.

In barley (*Hordeum vulgare* L.), restriction fragment length polymorphisms (RFLPs) (Graner et al., 1994) as well as amplified fragment length polymorphisms (AFLPs) (Schut et al., 1997) or simple sequence repeats (SSRs) (Russell et al., 1997) were successfully used to separate spring from winter cultivars. In wheat (*Triticum aestivum* L.) RFLP and SSR markers were applied to divide germplasm into high and low quality types (Bohn et al., 1999) or to evaluate close pedigree relationships (Plaschke et al., 1995). In addition, all studies yielded significant correlations between the coancestry coefficient (*f*) (Malécot, 1948) and GDs and were consequently useful for revealing pedigree relationships among cultivars.

As shown in all other major crops, molecular markers proved to be an important tool for studying genetic relationships of inbred lines within and between gene pools in maize (*Zea mays* L.). GDs based on molecular marker data were used to evaluate genetic diversity and to divide germplasm into gene pools. It was shown that GD values revealed by RFLPs (Messmer et al., 1993) as well as random amplified polymorphic DNA (RAPDs) (Hahn et al., 1995), AFLPs (Lübberstedt et al., 2000) or SSRs (Smith et al., 1997) are adequate estimates for the underlying pedigree relationships. Furthermore, Lübberstedt et al. (2000) and Pejic et al. (1998) demonstrated that GDs revealed by different marker systems were significantly correlated among each other and with *f* .

Hypothetically, *f* itself could be used for identification of EDVs, because it yields an estimator of the relationship between two genotypes and reflects the breeding procedure applied to derive a potential EDV from an IV. However, use of *f* is associated with several

major drawbacks, as it is only the expected value of the genetic similarity (1-GD) of two individuals and provides no information on the true genetic relatedness (Messmer et al., 1993). In addition, *f* is based on assumptions that are mostly unrealistic (Melchinger et al., 1991), *e.g.*, all lines in the pedigree pathway are homogeneous and homozygous, lines with no common parentage have *f*=0, or lines derived from a cross obtained half of the genome from each parent. Moreover, reliable pedigree data will most likely be unavailable in case of a suspected EDV.

The estimation of the conformity of the essential characteristics by the use of phenotypic data, such as morphological traits or heterosis, has been proposed and is still under debate (ASSINSEL, 1999a). Hitherto, accurate morphological and agronomic descriptions of cultivars and varieties are the basis of world-wide PVP systems and assure farmers and breeders that they are using clearly identifiable varieties to high standards of purity and quality (Smith and Smith, 1989a). In addition, numerous studies showed significant correlations between midparent heterosis and the coefficient of parentage (*f*) (Melchinger, 1999; Smith et al., 1991). For these reasons, proponents of morphological traits or heterosis for identifications of EDVs argue that phenotypic information provides the basis for PVP and should, thus, also be used for identification of EDVs.

Critical data on the scientific reliability for all kinds of distance measures, as requested by ASSINSEL (1999b) is still lacking. Potential reproducibility problems of molecular marker data were investigated by Jones et al. (1997), who reported scoring differences of up to 2 base pairs (bp) among the same SSR fragments detected by different laboratories. In sugar beet (*Beta vulgaris L.*) and wild *Beta* species, a reproducibility of AFLP bands of 97.6 % was determined by performing all necessary analytical steps twice (Hansen et al., 1999). In contrast to these results, Jones et al. (1997) and Bagley et al. (2001) reported an extremely high reproducibility of AFLP bands close to 100%. In addition, the relationship between morphological distances and GDs based on molecular markers or the coancestry coefficient (*f*) in maize was observed to be triangular instead of linear (Dillmann and Guérin, 1998). Furthermore, associations among maize inbred lines on the basis of morphology were not correlated with distance measures derived from heterosis or pedigree data (Smith and Smith, 1989b). However, no detailed comparison of GDs based on molecular markers, morphological data, and heterosis on their ability to accurately identify EDVs is available hitherto.

Objectives

The main goal of this study was to examine the usefulness of molecular markers, morphological traits, and heterosis for identifying EDVs in maize. In detail, the objectives of this thesis were to (1) investigate genetic and technical sources of variation in SSR and AFLP data within maize inbred lines and assess their impact on identification of EDVs, (2) analyze factors influencing GDs based on SSRs and AFLPs between related maize inbred lines, (3) investigate the power of SSR- and AFLP-based GD estimates for discriminating between progenies derived from F_2, BC_1, and BC_2 populations, (4) exemplify the theoretical and simulated results of the statistical theory introduced by Bohn et al. (2003) with experimental data, (5) compare the results observed for molecular markers to those obtained with morphological traits and heterosis, and (6) draw conclusions with regard to a detailed comparison of the various distance measures and EDV thresholds suggested in the literature.

References

ASSINSEL. 1999a. Essential derivation and dependence. Practical Information.

ASSINSEL. 1999b. Consolidation of ASSINSEL position papers on protection of biotechnological inventions and plant varieties.

Bagley, M. J., S. L. Anderson, and B. May. 2001. Choice of methodology for assessing genetic impacts of environmental stressors: polymorphism and reproducibility of RAPD and AFLP fingerprints. Ecotoxicology. 10:239-244.

Becker, H. C. Pflanzenzüchtung (in German). UTB, Stuttgart.

Bernardo, R. and A. L. Kahler. 2001. North American study on essential derivation in maize: inbreds developed without and with selection from F_2 populations. Theor. Appl. Genet. 102:986-992.

Bohn, M., H. F. Utz, and A. E. Melchinger. 1999. Genetic similarities among winter wheat cultivars determined on the basis of RFLPs, AFLPs, and SSRs and their use for predicting progeny variance. Crop Sci. 39:228-237.

Bohn, M., M. Frisch, M. Heckenberger, and A. E. Melchinger. 2004. A statistical framework for identification of essentially derived varieties obtained from biparental crosses of homozygous lines. Statistical theory. In Preparation.

Daily, G., P. Dasgupta, B. Bolin, P. Crosson, J. d. Guerny, P. Ehrlich, C. Folke, A. M. Jansson, B.-O. Jansson, N. Kautsky, A. Kinzig, S. Levin, K.-G. Mäler, P. Pinstrup-Andersen, D. Siniscalco, and B. Walker. 1998. Global food supply: food production, population growth, and the environment. Science. 281:1291-1292.

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

Duvick, D. N. and K. G. Cassman. 1999. Post-green revolution trends in yield potential of temperate maize in the north-central United States. Crop Sci. 39:1622-1630.

Gill, P., C. P. Kimpton, A. Urquhart, N. Oldroyd, E. S. Millican, S. K. Watson, and T. J. Downes. 1995. Automated short tandem repeat (STR) analysis in forensic casework: a strategy for the future. Electrophoresis. 16:1543-1552.

Gilliland, T. J., R. Coll, E. Calsyn, M. de. Loose, M. J. T. v. Eijk, and I. Roldan Ruiz. 2000. Estimating genetic conformity between related ryegrass (*Lolium*) varieties. 1. Morphology and biochemical characterisation. Mol. Breed. 6:569-580.

Graham, J., J. Curran, and B. S. Weir. 2000. Conditional genotypic probabilities for microsatellite loci. Genetics. 155:1973-1980.

Graner, A., W. F. Ludwig, and A. E. Melchinger. 1994. Relationships among european barley germplasm. 2. Comparison of RFLP and pedigree data. Crop Sci. 34:1199-1205.

Hahn, V., K. Blankenhorn, M. Schwall, and A. E. Melchinger. 1995. Relationships among early European maize inbreds. 3. Genetic diversity revealed with RAPD markers and comparison with RFLP and pedigree data. Maydica. 40:299-310.

Hansen, M., T. Kraft, M. Christiansson, and N. O. Nilsson. 1999. Evaluation of AFLP in *beta*. Theor. Appl. Genet. 98:845-852.

Heisey, P. W., C. S. Srinivasan, and C. Thirtle. 1999. Public sector plant breeding in a privatizing world. Agric. Inf. Bull. No. 772 US Department of Agriculture.

International Seed Federation. 2002. ISF view on Intellectual Property.

International Seed Federation. 2003. 2002/2003 Progress Report of the Secretariat.

Jones, C. J., K. J. Edwards, S. Castaglione, M. O. Winfield, F. Sala, C. Vandewiel, G. Bredemeijer, B. Vosman, M. Matthes, A. Daly, R. Brettschneider, P. Bettini, M. Buiatti, E. Maestri, A. Malcevschi, N. Marmiroli, R. Aert, G. Volckaert, J. Rueda, R. Linacero, A. Vazquez, and A. Karp. 1997. Reproducibility testing of RAPD, RFLP and SSR markers in plants by a network of european laboratories. Mol. Breed. 3:381-390.

Knaak, C., J. Förster, and Jäger-Gussen M. 1996. "Abgeleitete" Sorten aus praktischer Sicht (in German). Bericht über die 47. Arbeitstagung 1996 der Saatzuchtleiter im Rahmen der "Vereinigung österreichischer Pflanzenzüchter" BAL Gumpenstein. 167-172.

Lübberstedt, T., A. E. Melchinger, C. Dussle, M. Vuylsteke, and 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 Sci. 40:783-791.

Malécot, G. 1948. Les Mathematiques de l'Heredité. Masson & Cies, Paris.

Mann, C. 1997. BOTANY: Reseeding the Green Revolution. Science. 277:1038-1043.

Martin, J. M., T. K. Blake, and E. A. Hockett. 1991. Diversity among North American spring barley cultivars based on coefficients of parentage. Crop Sci. 31:1131-1137.

Melchinger, A.E. 1999. Genetic Diversity and Heterosis. Int. Symp. on Genetics and Exploitation of Heterosis in Crop Plants. Mexico City. p. 99-118. *In:* J.G. Coors and S. Pandey (*eds.*) The genetics and exploitation of heterosis in crops. ASA, CSSA, and SSSA, Madison, WI.

Melchinger, A. E., M. M. Messmer, M. Lee, W. L. Woodman, and K. R. Lamkey. 1991. Diversity and relationships among U.S. maize inbreds revealed by restriction fragment length polymorphism. Crop Sci. 31:669-678.

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

Pejic, I., P. Ajmone-Marsan, M. Morgante, V. Kozumplick, 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.

Plaschke, J., M. W. Ganal, and M. S. Roder. 1995. Detection of genetic diversity in closely related bread wheat using microsatellite markers. Theor. Appl. Genet. 91:1001-1007.

Reynolds, M. P., S. Rajaram, and K. D. Sayre. 1999. Physiological and genetic changes of irrigated wheat in the post-green revolution period and approaches for meeting projected global demand. Crop Sci. 39:1611-1621.

Roldan Ruiz, I., E. Calsyn, T. J. Gilliland, R. Coll, M. J. T. van Eijk, and De Loose M. 2000. Estimating genetic conformity between related ryegrass (*Lolium*) varieties. 2. AFLP characterization. Mol. Breed. 6:593-602.

Russell, J. R., J. D. Fuller, M. Macaulay, B. G. Hatz, A. Jahoor, W. Powell, and R. Waugh. 1997. Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. Theor. Appl. Genet. 95:714-722.

Schut, J. W., X. Qi, and P. Stam. 1997. Association between relationship measures based on AFLP markers, pedigree data and morphological traits in barley. Theor. Appl. Genet. 95:1161-1168.

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.

Smith J.S.C. and Smith O.S. 1989a. The description and assessment of distance between inbred lines of maize: I. The use of morphological traits as descriptors. Maydica. 34:141-150.

Smith J.S.C. and Smith O.S. 1989b. The description and assessment of distances between inbred lines of maize: II. The utility of morphological, biochemical, and genetic descriptors and a scheme for testing of distinctiveness between inbred lines. Maydica. 34:151-161.

Smith, J. S. C., O. S. Smith, S. L. Bowen, R. A. Tenborg, and 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 RFLPs. Maydica. 36: 213-226.

Tanksley, S. D. and S. R. McCouch. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. Science. 277:1063-1066.

Troyer, A. F. and T. R. Rocheford. 2002. Germplasm ownership: related corn inbreds. Crop Sci. 42:3-11.

UPOV. 1961. International convention for the protection of new varieties of plants.

UPOV. 1978. International convention for the protection of new varieties of plants.

UPOV. 1991. International convention for the protection of new varieties of plants.

Ustun, A., F. L. Allen, and B. C. English. 2001. Genetic progress in soybean of the U.S. midsouth. Crop Sci. 41:993-998.

Vasil, I. K. 1998. Biotechnology and food security for the 21st century: A real-world perspective. Nat. Biotech. 16:399-400.

Vaupel, J. W., J. R. Carey, K. Christensen, T. E. Johnson, A. I. Yashin, N. V. Holm, I. A. Iachine, V. Kannisto, A. A. Khazaeli, P. Liedo, V. D. Longo, Y. Zeng, K. G. Manton, and J. W. Curtsinger. 1998. Biodemographic trajectories of longevity. Science. 280:855-860.

Molecular Breeding **10:** 181–191, 2002.

Variation of DNA fingerprints among accessions within maize inbred lines and implications for identification of essentially derived varieties. I. Genetic and technical sources of variation in SSR data

Martin Heckenberger[1], Martin Bohn[1,*], Janet S. Ziegle[2], Larry K. Joe[2], Joan D. Hauser[2], Michelle Hutton[2] and Albrecht E. Melchinger[1]

*[1]Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany; [2]Celera AgGen, 1756 Picasso Avenue, Davis, CA 95616, USA; *Author for correspondence (e-mail: mbohn@uiuc.edu; fax: ++49 711 459 2343)*

Received 11 October 2001; accepted in revised form 19 March 2002

Key words: Essentially derived varieties, Genetic distances, Intra-varietal variation, Reproducibility, SSRs

Abstract

Genetic distances (GDs) based on molecular markers are important parameters for identifying essentially derived varieties (EDVs). In this context information about the variability of molecular markers within maize inbred lines is essential. Our objectives were to (1) determine the variation in the size of simple sequence repeat (SSR) fragments among different accessions of maize inbreds and doubled haploid (DH) lines, (2) attribute the observed variation to genetic and marker system-specific sources, and (3) investigate the effect of SSR fragment size differences within maize lines on the GD between maize lines and their consequences for the identification of essentially derived varieties. Two to five accessions from nine inbred lines and five DH lines were taken from different sources or drawn as independent samples from the same seed lot. Each accession was genotyped with 100 SSR markers that evenly covered the whole maize genome. In total, 437 SSR fragments were identified, with a mean of 4.4 alleles per locus. The average polymorphic information content (PIC) was 0.58. GD estimates between two accessions of the same genotype ranged from 0.00 to 0.12 with an average of 0.029 for inbred lines and 0.001 for DH lines. An average of 11.1 SSRs was polymorphic between accessions of the same inbred line due to non-amplification (8.1 SSRs), heterogeneity (4.0 SSRs) or unknown alleles (2.6 SSRs). In contrast to lab errors, heterogeneity contributed considerably to the observed variation for GD. In order to decrease the probability to be suited for infringing an EDV threshold by chance, we recommend to increase the level of homogeneity of inbred lines before applying for plant variety protection.

Introduction

According to the International Union for the Protection of new Varieties of Plants (UPOV) convention, a variety is deemed to be essentially derived from an initial variety if it is (i) predominantly derived and (ii) clearly distinguishable from the initial variety and (iii) genetically conform to the initial variety (UPOV 1991). The genetic conformity between initial and essentially derived varieties is considered to be the main important element within the concept of essentially derived varieties (EDV) (ASSINSEL 1999). Therefore, this concept should be based on genotypic information and the genetic distance (GD) between varieties is one of the key parameters to distinguish between essentially derived and independent varieties.

Genetic distances based on molecular marker data proved to be adequate estimates for the pedigree relationships in all major crops. Especially in maize, numerous studies yielded significant correlations between GDs obtained by molecular markers and the coefficient of coancestry (Lübberstedt et al. 2000; Smith et al. 1997). For this reason, molecular markers, particularly amplified fragment length polymorphisms (AFLPs) and simple sequence repeats (SSRs),

were proposed to be an appropriate tool to verify essential derivation in plant varieties (Smith et al. 1991; Knaak et al. 1996; ASSINSEL 2000).

Plant breeders did not yet implement the EDV concept into their breeding procedures due to the lack of suitable crop specific thresholds. EDV thresholds have to be crop specific to take into account the varying degree of polymorphism among the different crop species. For example, the proportion of alleles at marker loci alike in state between unrelated varieties is much smaller in maize than in barley or tomato (Qi and Lindhout 1997; Bernardo et al. 1997; Grandillo et al. 1999).

Because of the legal consequences scientifically reliable criteria have to be developed for the discrimination of essentially derived and independent varieties. Therefore, highest accuracy and reproducibility of GD estimates are mandatory. Potential reproducibility problems were investigated by Jones et al. (1997), who reported scoring differences of up to 2 bp among SSR fragments. In addition, information on the stability of molecular marker data over several generations of maintenance breeding is scanty.

The objectives of this study were to

- determine the variation in the size of SSR fragments among different accessions of maize inbreds and doubled haploid (DH) lines,
- attribute the observed variation to genetic and marker system-specific sources, and
- investigate the effect of SSR fragment size differences within maize lines on the GD between maize lines and their consequences for the identification of essentially derived varieties.

Material & methods

Plant materials

For nine maize inbred lines, six from the flint and three from the dent pool, and five DH lines from the dent pool, two to five accessions per line were fingerprinted (Table 1). Accessions were obtained from different generations of maintenance breeding conducted by the University of Hohenheim (UHOH) and three commercial breeding companies (B1–B3). Two accessions per DH line were obtained by drawing two independent samples of 20 kernels out of the same seed lot. The DH lines were derived from the cross of inbred lines RG2302 and 69117 by *in-vivo*-haploid induction (Deimling et al. 1997) with a subsequent colchicine treatment for chromosome doubling.

Maintenance breeding at the UHOH was performed by ear to row selection starting with the selfing of a single S_5 or S_6 ear per inbred line (Simmonds and Smartt 1999). In the further course of this procedure, one ear per row was selected per generation for maintaining the inbred line, whereas the other ears of the same homogeneous row were bulked for seed production. Thus, all individuals of one inbred line traced back to a single S_5 or S_6 ear.

All accessions of inbred lines were phenotypically homogeneous in field observation trials according to the regulations of the German Plant Variety Office and showed at least 95% of identical bands in an analysis of storage proteins conducted by isoelectric focusing following the rules of the International Seed Testing Association. Phenotypic evaluation was done at the UHOH experimental station at Eckartsweier, Germany, and the storage protein analyses were performed by the national agricultural research institute (LUFA) at Augustenberg, Germany, in 1999.

SSR analyses

DNA fingerprinting was performed with a standard marker set of 100 publicly available SSR markers that provide an even coverage of the maize genome (Figure 1). Fifteen seeds per accession were planted in a single pot for DNA extraction. Equal quantities of leaf material of 10 plants per accession were harvested at the 3 to 4 leaf stage. The leaf material was mixed and DNA was extracted using a modified Hexadecyltrimethyl-ammonium bromide (CTAB) procedure (Saghai Maroof et al. 1984). Electrophoresis was performed with an ABI Prism™ 377 DNA Sequencer using 5% polyacrylamid gels with 96 lanes. Internal fragment size standards were used in each lane to increase accuracy of DNA fragment size determination. The size of each DNA fragment was determined automatically by using the GeneScan® software and assigned to specific alleles by the Genotyper® software.

The 100 SSRs were selected based on robust single-locus amplification, absence of null alleles, high degree of polymorphism, and high reproducibility of the results. Seventy of the 100 SSRs contained dinucleotide repeat motifs, whereas the other 30 markers consisted of tri- to octa-nucleotide repeats. The development of the SSR set and the SSR analyses

Table 1. List of genotypes fingerprinted with 100 SSRs.

			Accessions		
Line	Type[†]	Pool	No.	Type[‡]	Source
D146	IL	Flint	3	1991, 1998, B3	bulk
D149	IL	Flint	4	1994, 1998, B1, B3	bulk
D171	IL	Flint	4	1994, 1998, B1, B3	bulk
D503	IL	Flint	2	1991, 1998	bulk
DK105	IL	Flint	4	1970 (1988)[§], 1980 (1988), 1991, 1996	bulk
UH002	IL	Flint	3	S_6, $S_{6:11}$, $S^{\#}_{6:11}$	ear
D06	IL	Dent	5	1988, 1994, 1998, B1, B2	bulk
UH200	IL	Dent	3	S_5, $S_{5:10}$, $S_{5:9}$	ear
UH300	IL	Dent	3	S_5, $S_{5:9}$, $S_{5:9}$	ear
RG2302[¶]	IL	Dent	1		
69117	IL	Dent	1		
941118[††]	HY	Dent	1		
ZS264	DH	Dent	2	2 repetitions[‡‡]	ear
ZS265	DH	Dent	2	2 repetitions	ear
ZS337	DH	Dent	2	2 repetitions	ear
ZS467	DH	Dent	2	2 repetitions	ear
ZS595	DH	Dent	2	2 repetitions	ear

[†] Line derivation: IL, inbred line; DH, doubled haploid line; HY, F_1 hybrid.
[‡] Accession types: year, year of seed maintenance performed by the UHOH maize program; B1–B3, line-maintenance performed by commercial breeders B1, B2, and B3.
[§] Maintenance breeding performed in 1970 and 1980 combined with one selfing generation in 1988 to maintain seed viability.
[#] Seeds of one S_5 or S_6 plant and two different S_9, S_{10}, or S_{11} plants derived from the particular S_5 or S_6 plant were used.
[¶] Parental lines of F_1 hybrid 941118.
[††] Parental source of all ZS lines.
[‡‡] Two independent samples from the second selfing generation of the same seed lot were fingerprinted.

were performed by Celera (1756 Picasso Avenue, Davis, CA 95616, USA).

Distinction of different cases of SSR results for various causes of variation

In order to distinguish different causes of variation for SSR results, the following distinction of disjunctive cases for each possible pairwise comparison of accessions of the same inbred or DH line was developed (Table 2).

Case 1 = The two accessions of a given line are homogeneous for the same allele A. This case is denoted as the normal case.

Case 2 = Both accessions are heterogeneous for the same alleles A and B.

Case 3 = Both accessions have the same allele A and one accession has an additional allele B.

Case 4 = One accession is homogeneous for allele A and one accession is heterogeneous carrying alleles B and C.

Case 5 = Both accessions are heterogeneous with only allele A in common.

Case 6 = Both accessions are heterogeneous with no allele in common.

Case 7 = One accession is heterogeneous and a missing value or null allele M occurs in the other accession.

Case 8 = One accession is homogeneous and a missing value or null allele M occurs in the other accession.

Case 9 = A missing value or null allele M occurs in both accessions.

Case 10 = The two accessions are homogeneous for different alleles.

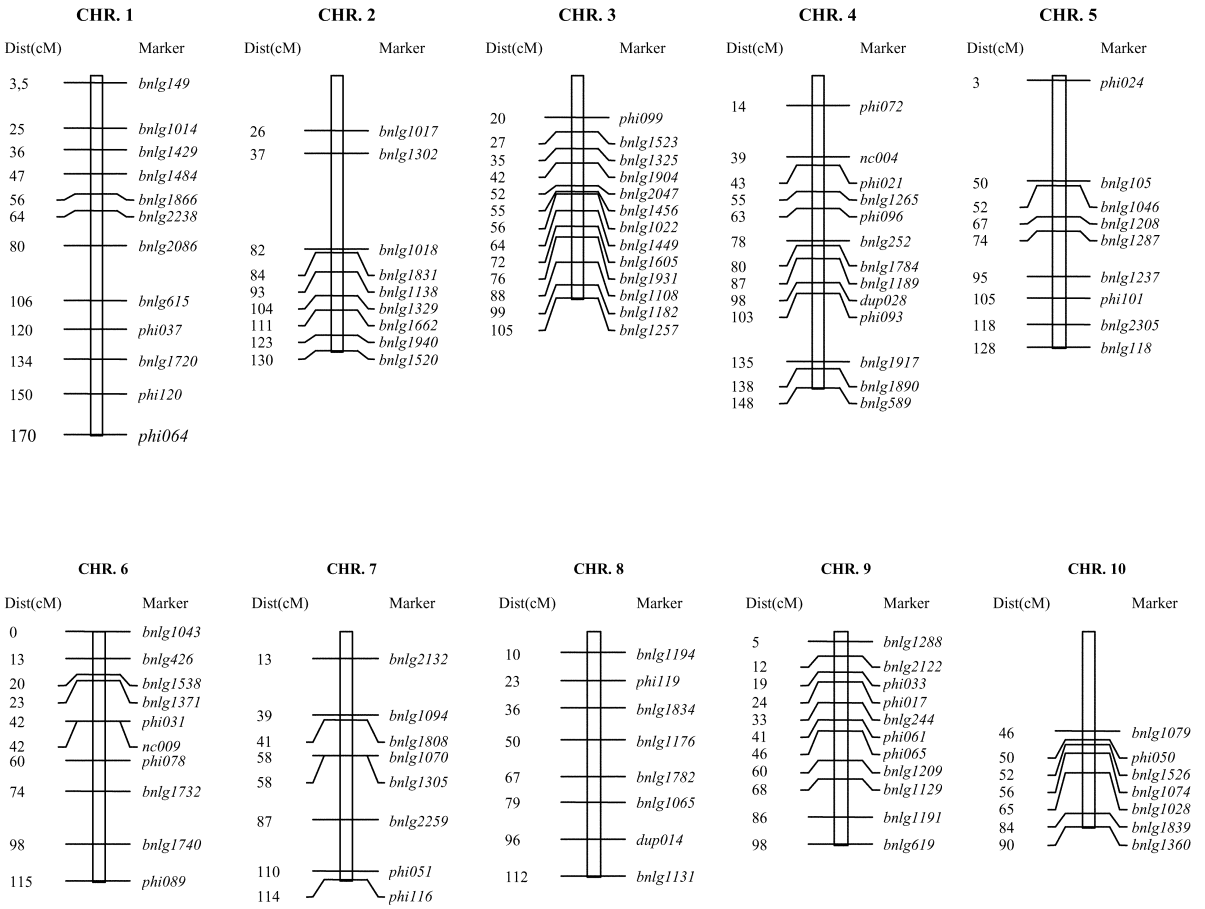


Figure 1. SSR markers used in the present study and their map positions.

Table 2. Distinction of different cases of SSR results for various causes of variation. Letters A, B, C, and D represent different alleles for a given marker locus, M represents a missing value. Accessions 1 and 2 denote two accessions of the same inbred or DH line.

	Case†									
Accession	1	2	3	4	5	6	7	8	9	10
1	A	A+B	A	A	A+B	A+B	A+B	A	M	A
2	A	A+B	A+B	B+C	A+C	C+D	M	M	M	B

† For a detailed description of the cases see Materials and Methods.

Statistical analyses

The polymorphic information content (PIC) was calculated for each SSR marker according to the formula of Botstein et al. (1980),

$$PIC = 1 - \left(\sum_{i=1}^{n} p_i^2 \right) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2 p_i^2 p_j^2,$$

where p_i and p_j are the frequencies of the i^{th} and j^{th} allele of a given marker, respectively. Genetic distances were calculated using the Dice coefficient (Nei and Li 1979). In the case of missing values, *i.e.*, one or several primer pairs did not yield an amplification product in one accession, the corresponding alleles of the other accession were not used for GD calculation. The cluster analysis was performed with the distance matrix using the UPGMA method (Nei et al. 1983). The reliability of the cluster was assessed by applying a bootstrap procedure (Efron 1979).

The fit of observed heterogeneity with the values from expected heterozygosity in S_5 or S_6 and S_9, S_{10}, or S_{11} generations was evaluated with a χ^2 test. Heterogeneity was defined as the number of marker loci that were not homozygous. Differences between dinucleotide repeats and SSRs with larger repeat motifs with regard to PIC values, number of alleles per marker, level of heterogeneity, number of fragment

size differences, and number of null alleles were tested with a *t*-test.

PIC values were calculated using the Cervus computer program (Marshall et al. 1998), estimation of GDs and cluster analysis were performed with the NTSYS-PC software package (Rohlf 1989). The bootstrap procedure was carried out with the Winboot computer program (Yap and Nelson 1996). The SAS software package was used for all other statistical calculations (SAS Institute 1988).

Results

Characterization of markers

A total of 437 SSR alleles were identified. The number of alleles per marker varied from 1 to 9 with an average of 4.4 alleles per marker. Only marker *bnlg1605* on chromosome 3 was monomorphic across all accessions. PIC values of polymorphic markers varied from 0.25 to 0.82 with an average of 0.58.

Genetic relationships of accessions within maize lines

The GD between two accessions of the same line ranged between 0.00 and 0.12 for inbred lines and between 0.00 and 0.01 for DH lines. The mean GD between accessions of the same line varied from 0.01 to 0.08 for inbred lines (Table 3). The average of the GD means was 0.03 for inbred lines and 0.00 for DH lines. The dendrograms obtained from UPGMA cluster analyses on the basis of GD estimates resulted in a clear separation of flint and dent inbred lines (Figure 2).

Sources of variation

Case 1: Normal case

For different accessions of a given inbred line, the average number of markers displaying the normal case varied from 78.7 for DK105 with 6 pairwise comparisons to 95.0 for D503 with 1 pairwise comparison (Table 4). Across the sets of DH lines, an average number of 92.0 loci showed the normal case, ranging from 87 for ZS265 to 98 for ZS337. Thirty-five SSR markers showed the normal case across all pairwise comparisons of accessions within all inbred or DH lines.

Table 3. Means, minima, and maxima of genetic distances (GD) between accessions of the same inbred line.

Line	No. of accessions	GD between accessions		
		Mean	Min	Max
D146	3	0.09	0.07	0.12
D149	4	0.03	0.04	0.02
D171	4	0.02	0.01	0.03
D503	2	0.01	0.01	0.01
DK105	4	0.04	0.02	0.06
UH002	3	0.03	0.03	0.04
D06	5	0.01	0.00	0.03
UH200	3	0.03	0.03	0.05
UH300	3	0.03	0.02	0.05
Total/Mean	31	0.029		

Cases 2–7: Heterogeneity

The number of marker loci heterogeneous for at least one accession per inbred or DH line varied from 1 to 13 for inbred lines and from 2 to 4 for DH lines. The observed level of heterogeneity for S_5 and S_6 lines was not significantly different from the expected heterozygosity levels. For S_9, S_{10}, and S_{11} lines, the observed level of heterogeneity was significantly ($P < 0.05$) higher than expected and was not significantly different from the observed heterogeneity for S_5 and S_6 lines. An average of 3.1 and 0.2 heterogeneous marker loci (Cases 3–7) was found for inbred and DH lines, respectively (Table 4). An average of 1.0 marker loci for inbred lines and 2.2 for DH lines showed a Case 2 type of heterogeneity with no effect on the variation of GD estimates. Four loci showed three DNA fragments for one accession at a particular locus (included in Cases 3–5).

Accessions of DH lines showed a GD of 0.0 with one exception. One accession of line ZS 264 carried allele 236 at marker locus *bnlg2122* whereas the second accession showed alleles 236 and 254.

Cases 7–9: Occurrence of null alleles or missing values

Forty-one markers showed no amplification product for at least one of all 44 fingerprinted accessions. Across all accessions of one specific inbred line, a mean of 10.3 SSRs yielded no amplification product. Across all DH lines, the mean number of markers without amplification product was 5.6 (data not shown). In addition, several cases of non-amplification in only one of the two repetitions of a DH line were detected.

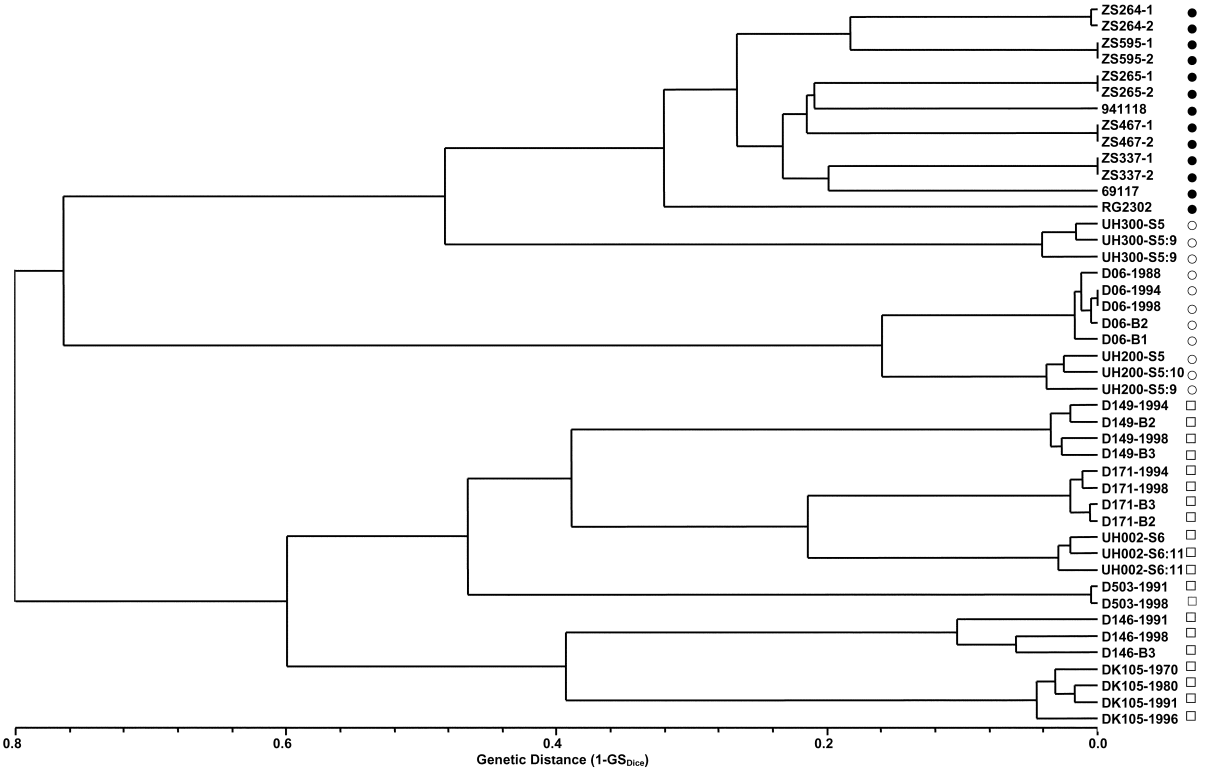


Figure 2. Associations among accessions of maize inbred lines revealed by UPGMA cluster analysis based on genetic distances calculated from SSR data. Asterisks (*) at the forks indicate that the group right of the fork was found for at least 95% of 1000 bootstrap runs. DH lines are marked by filled circles (●). Flint and dent lines are marked with squares (□) and circles (● ○), respectively.

Case 10: New alleles

For inbred lines the number of loci with homozygous unknown alleles averaged 3.9. Homozygous unknown alleles between accessions of the same line were not detected for DH lines (Table 5). The differences between fragment sizes of accessions of the same inbred lines was 1 bp in 14 cases and larger than 1 bp in 25 cases. For six of the 14 differences scored to 1 bp, the exact software-detected difference was 0.4 bp and smaller. The exact difference was 0.9–1.2 bp for the other eight 1 bp differences. The difference between exact and rounded values for allele calling was at 840 out of the 4561 data points 0.6 bp and higher. At 14% of the marker data points the difference was 0.6–0.8 bp, and between 0.8 and 1.2 bp at 4.3% of the data points. Unknown alleles increased the variation of GD within accessions of the same line. The fragment sizes of all accessions of a given line were rounded only in a few cases in the same wrong direction with no effect on the variation of GD estimates within the line. In addition, non-parental bands were detected at six marker loci for DH lines.

Genetic distance between lines

Due to the fact that 2 to 5 accessions per line were fingerprinted, GD values of up to 20 pairwise comparisons of accessions between the same two lines were calculated. The range for GD values calculated for each pairwise combination of accessions between the same two lines varied from 0.00 to 0.07 with an average range of 0.02. The mean range of GD between two inbred lines was significantly higher *($P < 0.01$)* than the mean range of GD between two DH lines.

Comparison of di-nucleotide and longer repeats

Di-repeat SSRs yielded a significantly greater number of alleles per marker ($P < 0.01$) and number of unknown alleles ($P < 0.05$) than SSRs with longer repeat motifs. The observed differences between these two SSRs groups were not significant for PIC values, the number of missing values, and the amount of heterogeneity.

Table 4. Occurrence of different cases for each pairwise comparison of accessions of the same line analyzed in this study.

Line	No. of pairwise comparisons	Case									
		1	2	3	4	5	6	7	8	9	10
		Average no. of loci per pairwise comparison									
Inbred lines											
D146	3	78.7	1.3	5.0	0.7	1.3	0.0	1.0	3.7	4.7	3.7
D149	6	84.7	1.8	4.0	0.0	0.0	0.0	0.3	5.2	2.0	2.0
D171	6	87.3	1.0	0.7	0.0	0.0	0.0	1.3	5.3	3.2	1.2
D503	1	95.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	4.0	0.0
DK105	6	78.7	0.7	1.7	0.0	1.0	0.0	0.5	9.0	6.0	2.5
UH002	3	88.0	1.0	3.7	0.0	0.0	0.0	0.3	2.3	4.0	0.7
D06	10	91.5	1.1	1.0	0.0	0.0	0.0	0.0	1.6	4.0	0.8
UH200	3	90.3	1.0	2.7	0.0	0.0	0.0	0.0	2.0	2.0	2.0
UH300	3	85.3	0.7	1.3	0.0	0.0	0.0	0.0	6.0	4.0	2.7
Mean		**86.6**	**1.0**	**2.3**	**0.1**	**0.3**	**0.0**	**0.4**	**3.9**	**3.8**	**1.7**
DH-lines											
ZS264	1	91.0	3.0	1.0	0.0	0.0	0.0	0.0	0.0	5.0	0.0
ZS265	1	87.0	2.0	0.0	0.0	0.0	0.0	0.0	9.0	2.0	0.0
ZS337	1	98.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ZS467	1	92.0	2.0	0.0	0.0	0.0	0.0	0.0	3.0	3.0	0.0
ZS595	1	92.0	2.0	0.0	0.0	0.0	0.0	0.0	2.0	4.0	0.0
Mean		**92.0**	**2.2**	**0.2**	**0.0**	**0.0**	**0.0**	**0.0**	**2.8**	**2.8**	**0.0**
Total Mean		**88.5**	**1.4**	**1.6**	**0.1**	**0.2**	**0.0**	**0.3**	**3.5**	**3.4**	**1.1**

Table 5. Number and category of differences in fragment size at the same locus within the same line group for Case 10.

Line	No. of accessions	Fragment size differences				Total
		1 bp	2 bp	3 bp	⩾4 bp	
		No. of differences				
D146	3	6	1	1	4	12
D149	4	1	2	0	2	5
D171	4	0	0	1	1	2
D503	2	0	0	0	0	0
DK105	4	5	1	0	2	8
UH002	3	0	1	0	1	2
D06	5	0	1	0	1	2
UH200	3	1	1	0	2	4
UH300	3	1	0	0	3	4
DH-lines	10	0	0	0	0	0
Total	41	14	7	2	16	39

Discussion

The use of flint and dent lines as well as the commonly employed methods of maintenance breeding make our study representative for maize breeding in Europe. In addition, SSR analyses were performed using publicly available SSR primers with a semi-automatic high-throughput system, which reduces human errors and subjectivity to a minimum. This should result in a higher accuracy of GD estimates, a prerequisite for EDV identification.

The degree of polymorphism in our line set was in close agreement with results reported by Smith et al. (1997). PIC values and the average number of alleles per marker were of similar size in both studies. However, in an analysis of genetic diversity among 33 inbred lines from the US corn belt, Pejic et al. (1998) found a substantially higher number of alleles than in the previous studies. The observed discrepancy can be explained by sampling effects caused by different inbred line sets that were fingerprinted with different sets of SSR markers.

Causes of variation

We observed a considerable variation for GD among different accessions of the same inbred line. This variation can be explained by genetic and technical reasons.

Genetic reasons

Mutations within the SSR primer region may yield null alleles, whereas a mutation between the primer regions may result in new alleles. The natural mutation rate for genomic non-repetitive DNA is estimated to range from 10^{-8} to 10^{-6} per locus and generation (Drake et al. 1998; Allard 1999). However, SSRs showed higher mutation rates than non-SSR regions ranging from approximately 10^{-6} per locus and generation for *Saccharomyces cerevisiae* (Sia et al. 2000) up to 10^{-3} in the pipefish *Syngnathus typhle* (Jones et al. 1999). The mutation rate of SSRs was found to be dependent on the repeat type, the repeat number, and the sequence of the repeat motif or the flanking sequence (Schloetterer 2000). Mutations within SSR markers were mostly insertions and deletions of mainly complete repeats (Twerdi et al. 1999). For maize, no information about the mutation rate of SSR loci is yet available. However, if their mutation rate is also higher than for non-SSR regions, mutations cannot be neglected as a cause of genetic variation between accessions of the same line. Unequal crossover in SSR regions is another genetic reason for the unexpected variation in GD as reported in wheat (Plaschke et al. 1995).

Segregation from S_5 and S_6 to the particular S_9, S_{10}, or S_{11} generations was the cause of genetic variation between accessions of the same line at four loci. At these loci, the fingerprinted S_5 or S_6 accession was heterogeneous and the corresponding S_9, S_{10}, or S_{11} accessions were homogeneous, each with one allele of the particular S_5 or S_6 accession. However, for S_9, S_{10}, and S_{11} lines the observed level of heterogeneity was significantly ($P < 0.05$) higher than expected for these selfing generations.

Bulking during maintenance breeding can be one reason for this unexpectedly high level of heterogeneity in highly inbred lines. Ears of each row not used for generating the next generation were bulked for seed production of the particular line. Because of the segregation due to residual heterozygosity, this procedure may have resulted in a mixture of genotypes that were homogeneous for different parental alleles. For samples drawn out of the bulk (Table 1), these effects are not negligible. In addition, genotyping a bulk of ten individuals can lead to variation, when certain regions of the genome still segregate. To avoid this, the individuals should be genotyped independently.

Another possible cause for the variation of GD values within the same line is contamination by foreign pollen during maintenance breeding (Smith et al. 1997). However, all accessions were homogeneous in field observation trials and isoelectric focusing. In addition, deviations from the normal case at numerous loci would be expected in case of a contamination. Therefore, contamination with foreign pollen can be excluded as a cause for the observed high level of heterogeneity in advanced generations.

Technical reasons

In this study, DH lines were used to distinguish between genetic and technical reasons for variation of GD within accessions of the same line. If the fingerprints of two samples of the same DH line are not identical, only technical reasons might explain this finding. The observation of heterozygous DH lines could be explained by heteroduplex bands, especially in those four cases with three alleles at a single locus. It is known that heteroduplex bands emerge by annealing of two DNA fragments of unequal sequence or length (Hatcher et al. 1993). This mismatch usually tends to retard the migration of DNA of the heteroduplex band during electrophoresis (Nataraj et al. 1999). In addition, the intensity of heteroduplex bands is supposed to fall between the corresponding homoduplex bands. However, adopting this definition, only two of the four above mentioned bands, would be heteroduplex.

Artificial stutter bands could also have led to variation in GD estimates within the inbred lines. Especially SSRs with a di-repeat motif are known to show stutter bands caused by a ‘loop’ of 2 bp in the strand of the template (Smith et al. 1997). Therefore, stutter bands appear 2 bp shorter than the main band. The software program “Genotyper” automatically identifies stutter bands based on their migration distance and the intensity of the particular bands. However, intense stutter bands could have been wrongly scored as non-stutter bands and independent bands with low intensity could have been incorrectly identified as stutter bands.

Compared to a study of Murigneux et al. (1993) performed with RFLPs, our results showed an increased level of heterogeneity between DH accessions based on SSRs. This finding can be explained by artificial SSR bands caused by heteroduplex or stutter bands. In addition, mutations due to colchicine treatment used for chromosome doubling and tissue culture steps may also have caused the observed heterogeneity (Marhic et al. 1998).

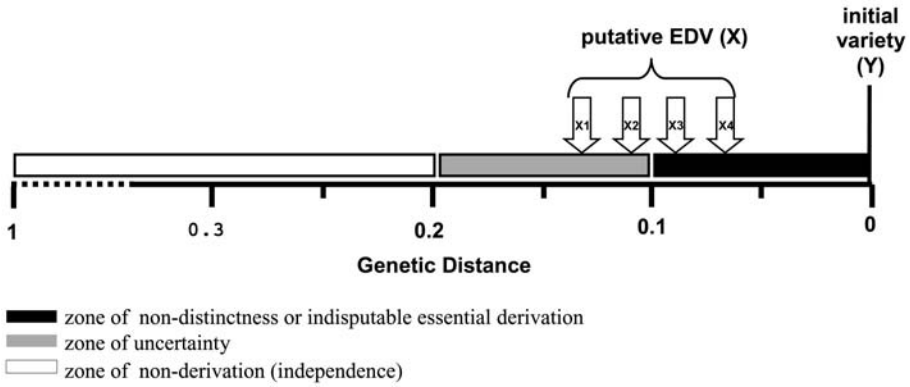


Figure 3. The concept of different thresholds and the consequences of variation between accessions of the same line for the identification of EDVs.

An additional source of variation within an inbred line are unexpected effects due to slippage of the DNA polymerase during the PCR. Therefore, whole or partial repeats could be added or removed from the template and thus yielding genotyping errors (Palsboll et al. 1999; Schloetterer and Tautz 1992).

Software imperfections resulted in a further increase of variation of GD estimates. Genotyping errors of the software Genotyper® caused about half of the 1 bp differences. The software Gene Scan® determined the DNA fragment size using a sizing curve with an accuracy of 0.1 bp. These fragment size measures were employed by the software Genotyper® to assign the fragments to specific alleles. However, reasons for the genotyping errors remain unknown as the algorithm of the software is not publicly available. For those 1 bp differences that could not be assigned to genotyping errors, slippage effects are a possible reason. In addition, small variations in the concentration of the gel, buffer or the voltage of the run could lead to 1 bp differences. To check the repeatability of these 1 bp differences, further studies are required, using several repetitions of single accessions, by sequencing the DNA of the particular fragments, or by genotyping individuals instead of bulks.

Implications for the identification of EDVs

The results of this study demonstrated that lab error and heterogeneity caused variation of GD estimates between different accessions of the same inbred line. In order to assess the implications of the lab error on the identification of EDVs, we first assume the following hypotheses:

H_0: Lines X and Y carry for a specific marker locus the same marker band.

H_1: Lines X and Y carry for a specific marker locus different marker bands.

Based on these hypotheses a Type I error, *i.e.*, two bands were scored as different although they were identical, and a Type II error, *i.e.*, two bands were scored as identical although they were different, can be distinguished. In the case of highly related lines, lab errors result in an overestimation of GD(X, Y) and the hypothesis that X and Y are highly related will be rejected too frequently. However, this study showed that for SSRs employing a semi-automated gel and scoring system, lab errors accounted only for a minor proportion of the detected variation of GD among accessions of the same inbred line and are, therefore, negligible.

The impact of heterogeneity on EDV identification can be exemplified using the following scenario. An individual of line X is used for the development of line Y. A genotypic fingerprint of line Y will be compared with fingerprints of different accessions of line X. Due to the possible varying GD values among accessions of X, GD(X,Y) will also vary conditional on the used accession of X. If the range of GD values between accessions of the same line is large, it might be possible that a true EDV could be judged as independently derived or a truly independent variety as essentially derived just by genotyping different accessions of each variety (Figure 3). In contrast to lab errors that were of negligible importance, heterogeneity reached considerable levels in some cases, *e.g.*, inbred line D146 (Table 3).

Our results illustrate the crucial importance of increased levels of homogeneity within new lines be-

fore applying for plant breeder's rights. If lines are heterogeneous, further selfing generations should be performed. A more rapid procedure would be the prescreening of newly developed lines with molecular markers. Our study showed that the variation of GD estimates within maize lines was lowest among DH lines. If DH techniques can efficiently be applied in a breeding program, this technique might be less laborious and costly than a routine fingerprinting with a marker set covering the entire genome.

Acknowledgements

The authors wish to thank Semundo (Hadmersleben, Germany) for supplying the DH lines.

References

Allard R.W. 1999. Principles of Plant Breeding. John Wiley & Sons, New York.

ASSINSEL 1999. Consolidation of ASSINSEL position papers on protection of biotechnological inventions and plant varieties.

ASSINSEL 2000. DUS testing: Phenotype vs. Genotype. ASSINSEL, Position Paper adopted at the Rome Congress in May 2000.

Bernardo R., Murigneux A., Maisonneuve J.P., Johnsson C. and Karaman Z. 1997. RFLP-based estimates of parental contribution to F_2 and BC_1 derived maize inbreds. Theor. Appl. Genet. 94: 652–656.

Botstein D., White R.L., Skolnick M. and Davis R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32: 314–331.

Deimling S., Röber F. and Geiger H.H. 1997. Methodology and genetics of *in vivo* haploid induction in Maize. Vortr. Pflanzenzüchtg. 38: 203–224.

Drake J.W., Charlesworth B., Charlesworth D. and Crow J.F. 1998. Rates of spontaneous mutation. Genetics 148: 1667–1686.

Efron B. 1979. Bootstrap methods: another look at the jackknife. Ann. Stat. 7: 1–26.

Grandillo S., Ku H.M. and Tanksley S.D. 1999. Identifying the loci responsible for natural variation in fruit size and shape in tomato. Theor. Appl. Genet. 99: 978–987.

Hatcher S.L., Lambert Q.T., Raymond L.T. and Carlson J.R. 1993. Heteroduplex formation: a potential source of errors from PCR products. Prenatal Diagnosis 13: 171–177.

Jones A.G., Rosenqvist G., Berglund A. and Avise J.C. 1999. Clustered microsatellite mutations in the pipefish *Syngnathus typhle*. Genetics 152: 1057–1063.

Jones C.J., Edwards K.J., Castaglione S., Winfield M.O., Sala F., Vandewiel C. et al. 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Mol. Breed. 3: 381–390.

Knaak C., Förster J. and Jäger-Gussen M. 1996. "Abgeleitete" Sorten aus praktischer Sicht (in German). Bericht über die 47. Arbeitstagung 1996 der Saatzuchtleiter im Rahmen der "Vereinigung österreichischer Pflanzenzüchter" BAL Gumpenstein: 167–172.

Lübberstedt T., Melchinger A.E., Dussle C., Vuylsteke M. and Kuiper M. 2000. Relationships 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.

Marhic A., Anatoine-Michard S., Bordes J., Pollacsek M., Murigneux A. and Beckert M. 1998. Genetic improvement of anther culture response in maize: relationships with molecular, mendelian and agronomic traits. Theor. Appl. Genet. 97: 520–525.

Marshall T.C., Slate J., Kruuk L.E.B. and Pemberton J.M. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Mol. Ecol. 7: 639–655.

Murigneux A., Barloy D., Leroy P. and Beckert M. 1993. Molecular and morphological evaluation of doubled haploid lines in maize. 1. Homogeneity within DH lines. Theor. Appl. Genet. 86: 837–842.

Nataraj A.J., Olivos-Glander I., Kusukawa N. and Highsmith W.E. 1999. Single-strand conformation polymorphism and heteroduplex analysis for gel-based mutation detection. Electrophoresis 20: 1177–1185.

Nei M. and Li W.H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Soc. USA 76: 5269–5273.

Nei M., Tajima F. and Tateno Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. J. Mol. Evol. 19: 153–170.

Palsboll P.J., Berube M. and Jorgensen H. 1999. Multiple Levels of Single-Strand Slippage at Cetacean Tri- and Tetranucleotide Repeat Microsatellite Loci. Genetics 151: 285–296.

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

Plaschke J., Ganal M.W. and Roder M.S. 1995. Detection of genetic diversity in closely related bread wheat using microsatellite markers. Theor. Appl. Genet. 91: 1001–1007.

Qi X. and Lindhout P. 1997. Development of AFLP markers in barley. Mol. Gen. Genet. 254: 330–336.

Rohlf F.J. 1989. NTSYS-pc Numerical Taxonomy and Multivariate Analysis System. Exeter Publishing Co, Ltd, Setauket, NY.

Saghai Maroof M.A., Soliman K.M., Jorgensen R.A. and Allard R.W. 1984. Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. Proc. Natl. Acad. Soc. USA 81: 8014–8.

SAS Institute 1988. SAS/STAT User's Guide, Release 6.03 edn. SAS, Cary.

Schloetterer C. and Tautz D. 1992. Slippage synthesis of simple sequence DNA. Nucl. Acids Res. 20: 211–215.

Schloetterer C. 2000. Evolutionary dynamics of microsatellite DNA. Chromosoma 109: 365–371.

Sia E.A., Butler C.A., Dominska M., Greenwell P., Fox T.D. and Petes T.D. 2000. Analysis of microsatellite mutations in the mitochondrial DNA of *Saccharomyces cerevisiae*. Proc. Natl. Acad. Soc. USA 97: 250–255.

Simmonds N.W. and Smartt J. 1999. Principles of Crop Improvement. Blackwell Science, Oxford.

Smith J.S.C., Chin E.C.L., Shu H., Smith O.S., Wall S.J., Senior M.L. et al. 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.

Smith J.S.C., Smith O.S., Bowen S.L., Tenborg R.A. and Wall S.J. 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 RFLPs. Maydica 36: 213–226.

Twerdi C.D., Boyer J.C. and Farber R.A. 1999. Relative rates of insertion and deletion mutations in a microsatellite sequence in cultured cells. Proc. Natl. Acad. Soc. USA 96: 2875–2879.

UPOV 1991. International Convention for the Protection of New Varieties of Plants.

Yap I.V. and Nelson R.J. 1996. WinBoot: A Program for Performing Bootstrap Analysis of Binary Data to Determine Confidence Limits of UPGMA-Based Dendrograms. International Rice Research Institute, Manila, Philippines.

Molecular Breeding **12:** 97–106, 2003.

Variation of DNA fingerprints among accessions within maize inbred lines and implications for identification of essentially derived varieties: II. Genetic and technical sources of variation in AFLP data and comparison with SSR data

Martin Heckenberger[1], Jeroen Rouppe van der Voort[2,4], Albrecht E. Melchinger[1,4], Johan Peleman[2] and Martin Bohn[1,3,*]

*[1]Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany; [2]Keygene, P.O. Box 216, Wageningen, 6700, AE, The Netherlands; [3]Current address: Crop Science Department, University of Illinois, S-110 Turner Hall, 1102 South Goodwin Avenue, Urbana, IL 61801, USA; [4]These authors contributed equally to this work; *Author for correspondence (e-mail: mbohn@uiuc.edu; fax: ++217 333 9817)*

Received 14 May 2002; accepted in revised form 12 February 2003

Key words: AFLPs, Essentially derived varieties, Genetic distances, Intra-varietal variation, Reproducibility, SSRs

Abstract

Accuracy and reproducibility of genetic distances (GDs) based on molecular markers are crucial issues for identification of essentially derived varieties (EDVs). Our objectives were to investigate (1) the amount of variation for amplified fragment length polymorphism (AFLP) markers found among different accessions within maize inbreds and doubled haploid (DH) lines, (2) the proportion attributable to genetic and technical components and marker system specific sources, (3) its effect on GDs between maize lines and implications for identification of EDVs, and (4) the comparison to published SSR data from the same plant materials. Two to five accessions from nine inbred lines and five DH lines were taken from different sources of maintenance breeding or drawn as independent samples from the same seed lot. Each of the 41 accessions was genotyped with 20 AFLP primer combinations revealing 988 AFLP markers. Map positions were available for 605 AFLPs covering all maize chromosomes. On average, six (0.6%) AFLP bands were polymorphic between different accessions of the same line. GDs between two accessions of the same line averaged 0.013 for inbreds and 0.006 for DH lines. The correlation of GDs based on AFLPs and SSRs was tight (r = 0.97**) across all 946 pairs of accessions but decreased (r = 0.55**) for 43 pairs of accessions originating from the same line. On the basis of our results, we recommend specific EDV thresholds for marker systems with different degree of polymorphism. In addition, precautions should be taken to warrant a high level of homogeneity for DNA markers within maize lines before applying for plant variety protection.

Introduction

In all major crops, genetic distances (GDs) based on reliable molecular marker data have been found to reflect accurately the degree of pedigree relationships between genotypes (Melchinger 1999). In maize, several studies reported highly significant correlations between GDs based on molecular markers and the coefficient of coancestry (for review, see Lübberstedt et al. 2000). Consequently, molecular markers, especially amplified fragment length polymorphisms (AFLPs) and simple sequence repeats (SSRs), were recommended as an appropriate tool to test for essential derivation in plant varieties (Smith et al. 1991; Knaak et al. 1996; ASSINSEL 2000).

Scientifically reliable criteria must be developed to differentiate between EDVs and independent varieties because of the severe legal consequences for the

breeders. Indispensable prerequisites are accuracy and reproducibility of GD estimates. Reproducibility problems were investigated by Jones et al. (1997), who reported scoring differences of up to 2 base pairs (bp) among the same SSR fragments detected by different labs. In sugar beet (*Beta vulgaris* L.) and wild *Beta* species, a reproducibility of AFLP bands of 97.6% was determined by performing all necessary analytical steps twice (Hansen et al. 1999). In contrast to these results, Jones et al. (1997) and Bagley et al. (2001) reported an extremely high reproducibility of AFLP bands close to 100%. In addition, Heckenberger et al. (2002) revealed variation in GD estimates based on SSRs of up to 0.12 on a 0 to 1 scale between different generations of maintenance breeding of the same inbred line or the same inbred line maintained by different breeders. However, critical information on the reproducibility of AFLP bands and their stability during maintenance is still lacking.

The overall goal of our study was to determine the variation of AFLP markers among different accessions of maize inbreds and doubled haploid (DH) lines. In detail, our objectives were to investigate (1) the amount of variation for amplified fragment length polymorphism (AFLP) markers found among different accessions within maize inbreds and doubled haploid (DH) lines, (2) the proportion attributable to genetic and technical components and marker system specific sources, (3) its effect on the GD between maize lines and implications for identification of EDVs and (4) the comparison to published SSR data from the same plant materials (Heckenberger et al. 2002).

Materials and methods

Plant materials

For nine maize inbred lines, six from the flint and three from the dent pool, and five DH lines from the dent pool, two to five accessions per line were fingerprinted. Accessions were obtained from different generations of maintenance breeding conducted by the University of Hohenheim (UHOH) and three commercial breeding companies (B1–B3). Two accessions per DH line were obtained by drawing two independent samples of 20 kernels out of the same seed lot. All DH lines were derived from the cross 941118 of inbred lines RG2302 and 69117. A detailed description of the plant materials analyzed and the applied method of maintenance breeding is given in our companion paper (Heckenberger et al. 2002).

AFLP and SSR analyses

AFLP fingerprints were generated by Keygene N.V. from leaf punches from a bulk of 10 individual plants per accession as described by Vos et al. (1995), visualized by use of a Fuji BAS/2000 phosphorimager, and scored dominantly on the set of maize lines with proprietary software developed by Keygene N.V. AFLP markers were referred to a proprietary integrated map of maize. This integrated map combines linkage information of 5650 molecular markers and is based on 23 separate mapping populations (Peleman et al. 2000). Comparison of the 988 AFLP markers scored in this project with the integrated maize map revealed that for 605 of these AFLP markers map information was available.

SSR analyses were carried out by Celera using 100 publicly available SSR primer pairs, equally distributed across the maize genome using an ABI Prism™ 377 DNA Sequencer with 5% polyacrylamide gels. Internal fragment size standards were used in each lane to increase accuracy of DNA fragment size determination. Fragment sizes were determined automatically by using the GeneScan® and Genotyper® software packages. For a detailed description of the procedures, see our companion paper (Heckenberger et al. 2002). AFLP and SSR analyses of each accession were carried out using seeds from the same seed lot.

Statistical analyses

The polymorphic information content (PIC) was calculated for each primer combination using the formula $PIC = 2p_i\,(1 - p_i)$, where p_i is the frequency of the i^{th} AFLP band (Roldan-Ruiz et al. 2000). The marker index was calculated for each AFLP primer combination as $MI = \overline{PIC} \times n\beta$, where $\overline{PIC}$ is the mean PIC value, n is the number of bands, and β is the proportion of polymorphic bands (Powell et al. 1996).

Genetic distances (GDs) were calculated as 1 – genetic similarity (GS). For AFLPs, GS was calculated using the GS coefficient of Jaccard (1908) and for SSRs the GS coefficient of Dice (1945) was employed. In the case of missing values, *i.e.*, if one or several primer combinations did not yield amplification products in one of the two accessions compared,

the corresponding alleles of the other accession were not used for GD calculation. GD estimates were based on the whole AFLP data set or on AFLP data for single chromosomes. The cluster analysis was performed with GD estimates using the 'unweighted pair group method using arithmetic averages' (UPGMA) (Nei et al. 1983). The reliability of the cluster was assessed by applying a bootstrap procedure (Efron 1979). The cophenetic correlation (r_{cpe}) was calculated to test for the goodness-of-fit between GD values obtained from the cluster and the original GD estimates. The significance of r_{cpe} was determined by the Mantel test (Mantel 1967) based on 10000 permutations. Standard deviation (SD) of GD estimates was calculated using the formula of Bar-Hen and Charcosset (1994),

$$SD = \sqrt{\frac{1}{N} GD \times (1 - GD)},$$

where N is the number of polymorphic bands and GD is the genetic distance between two genotypes or the mean GD between two groups of genotypes.

The null hypothesis (H_0) that the markers were randomly distributed across the genome was tested against the alternative hypothesis (H_1) that the markers were not randomly distributed across the genome by a dispersion analysis (Johnson et al. 1992). The test statistic was

$$X^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{\bar{y}}.$$

Here, n denotes the number of chromosomal intervals between two fixed marker loci (BINs), y_i the number of markers located in a particular BIN, and $\bar{y}$ the mean number of markers per BIN. Under H_0, X^2 follows a χ^2-distribution with n-1 degrees of freedom. H_0 was rejected, if X^2 was smaller than $\chi^2_{1-\frac{\alpha}{2};n-1}$ or larger than $\chi^2_{\frac{\alpha}{2};n-1}$.

GDs between accessions of the same line were defined as ″GD within″ (GD_w). Lines were screened for outliers based on their mean GD_w values, by applying the test of Anscombe and Tukey (1963). For all analyses where information on map positions was necessary, only the mapped AFLP markers were used. The genomic locations of variation within maize lines detected with AFLPs and SSRs were compared based on their BIN positions.

Calculation of GDs, cluster analysis, cophenetic correlation and Mantel test were performed with software NTSYS-PC (Rohlf 1989). The bootstrap procedure was carried out with the Winboot computer program (Yap et al. 1996). All other statistical calculations were performed with SAS (SAS Institute 1988).

Results

Description of markers

A total of 988 AFLP bands was identified. The number of polymorphic bands per PC varied from 40 to 83 with an average of 49. No significant differences between *Pst*I/*Mse*I and *Eco*RI/*Mse*I primers were found for the number of polymorphic bands, mean *PIC*, *MI*, and mean GD_w values (Table 1). In addition, the proportion of missing values had no influence on these quality parameters. Comparison of the 988 AFLP markers scored in this project with the Keygene integrated maize map (Peleman et al. 2000) revealed that for 605 of these AFLP markers map information was available. The average marker interval of the mapped markers was 2.5 cM, the total map length amounted to 1512 cM.

By comparison, the average marker interval for SSRs employed in the companion study was 12.4 cM and the total map length was 1210 cM (Heckenberger et al. 2002). Evaluation of the number of markers per BIN revealed a higher standard deviation for AFLPs than for SSRs. The null hypothesis of a random marker distribution across the genome was rejected for both AFLPs and SSRs. For AFLPs, X^2 was significantly ($P < 0.05$) larger than $\chi^2_{\frac{\alpha}{2};n-1}$, indicating that AFLPs were heterogeneously distributed over the genome, whereas X^2 was smaller than $\chi^2_{1-\frac{\alpha}{2};n-1}$, indicating that SSRs were uniformly distributed across the genome.

Genetic distances of accessions within maize lines

Between different accessions of the same inbred or DH line, an average of six AFLP fragments (0.6%) was polymorphic. The mean GD_w for inbred lines calculated separately for each line varied from 0.000 to 0.022 (Table 2).

Table 1. Statistics characterizing the degree of polymorphism and quality of AFLP data generated with 20 primer combinations

	Enzymes		Polymorphic bands					
PC[†]	*Eco*RI/*Mse*I	*Pst*I/*Mse*I	No.	Proportion (%)	$\overline{PIC}$ [‡]	MI[§]	$\overline{GD_w}$ [#]	Prop. Of missing values (%)[¶]
A	x		45	66	0.36	16.5	0.010	2
B		x	46	73	0.36	16.6	0.010	6
C	x		40	47	0.38	15.2	0.005	2
D	x		58	69	0.34	19.8	0.023	6
E		x	40	58	0.35	13.9	0.012	6
F		x	56	68	0.26	14.7	0.024	28
G		x	47	62	0.30	14.0	0.019	9
H		x	41	59	0.37	15.2	0.019	5
I	x		59	68	0.36	21.2	0.009	8
J		x	46	69	0.35	16.1	0.007	3
K	x		43	58	0.35	15.3	0.030	6
L	x		83	71	0.34	27.8	0.006	4
M		x	45	62	0.38	17.0	0.012	4
N	x		51	68	0.36	18.4	0.012	3
O	x		48	67	0.34	16.3	0.016	9
P		x	46	70	0.34	15.5	0.023	5
Q	x		48	62	0.36	17.4	0.016	2
R		x	50	68	0.35	17.6	0.018	6
S	x		47	64	0.27	12.5	0.006	38
T		x	49	68	0.35	17.2	0.013	7
Mean			**49**	**65**	**0.34**	**16.5**	**0.015**	**8**

[†] PC = Primer combination.
[‡] $\overline{PIC}$ = Mean PIC value observed for AFLPs of the particular PC.
[§] MI = Marker index.
[#] $\overline{GD_w}$ = Mean GD value between accessions of the same inbred or DH line (GD_w), obtained only with markers from the particular PC.
[¶] Proportion of missing values based on all datapoints.

The dendrogram obtained from UPGMA cluster analysis resulted in a clear separation of flint and dent inbred lines (Figure 1). Using genotypic data from all chromosomes, accessions derived from the same line were always clustered together. However, this pattern was not consistently obtained comparing the 10 dendrograms based on marker data of individual chromosomes (data not shown).

GD estimates based on AFLP and SSR data were significantly ($P < 0.01$) correlated with each other (r = 0.97**) for all pairwise 946 combinations of all 41 entries (Table 3). The correlation dropped to r = 0.07 when considering only the 46 GD_w values between accessions of the same line and increased to r = 0.55** after omitting the GD_w values of outliers ZS 467 (AFLP) and D146 (SSR) (Figure 2).

Locations of variation within maize lines

A total of 58 mapped AFLP markers, localized within 31 BINs, showed variation between accessions of the same line. Only six of these BIN locations displayed also variation with SSRs for the same accessions. These were located on chromosomes 1 (DK105), 2 (D149, D171, UH200), 7 (D149), and 8 (UH002). For all other BINs with varying AFLP markers adjacent SSR markers were not polymorphic within lines.

Genetic distance between lines

The effect of the variation of GD_w on the GD between lines was investigated by calculating the span for GD values for each pairwise combination of accessions between the same two lines. These span values varied from 0.001 to 0.114 with an average of 0.017. If outliers D146 and ZS467 were omitted from the anal-

Table 2. Mean, maximum, and minimum of genetic distances (GD_w) between accessions of the same maize inbred or DH line

		AFLP			SSR		
Line	No. of accessions	Mean	Min.	Max.	Mean	Min.	Max.
Inbred lines							
D146	3	0.004	0.000	0.006	***0.090*** †	0.061	0.116
D149	4	0.020	0.008	0.033	0.031	0.020	0.042
D171	4	0.004	0.000	0.008	0.016	0.005	0.028
D503	2	0.000	–‡	–‡	0.005	–‡	–‡
DK105	4	0.022	0.008	0.032	0.036	0.017	0.056
UH002	3	0.009	0.005	0.016	0.026	0.021	0.031
D06	5	0.011	0.003	0.020	0.011	0.000	0.021
UH200	3	0.013	0.010	0.015	0.034	0.025	0.047
UH300	3	0.020	0.019	0.022	0.033	0.016	0.044
Mean §		**0.011**			**0.024**		
DH lines							
ZS264	2	0.008	–‡	–‡	0.005	–‡	–‡
ZS265	2	0.008	–	–	0.000	–	–
ZS337	2	0.003	–	–	0.000	–	–
ZS467	2	***0.124*** †	–	–	0.000	–	–
ZS595	2	0.008	–	–	0.000	–	–
Mean §		**0.007**			**0.001**		
Grand mean §		**0.010**			**0.015**		

† Outliers based on the test of Anscombe and Tukey (1963).
‡ For lines with only two accessions, only a single GD_w value was available, therefore, no maximum and minimum was calculated.
§ Means were calculated leaving out outliers.

ysis, span-values for AFLPs and SSRs were of similar magnitude.

Discussion

AFLPs proved to be an appropriate tool to distinguish between flint and dent lines. This is in agreement with results published by Lübberstedt et al. (1999), who genotyped a set of 51 European flint and dent lines with eight AFLP primer combinations and found a clear separation of flint and dent germplasm. *PIC* values and marker indices were almost identical to those published by Lübberstedt et al. (2000). In addition, DH lines and their parental lines could be clearly separated from other dent germplasm developed by the University of Hohenheim. However, using only data from individual chromosomes, accessions of the same line did not cluster together in some cases. This lack of association can be explained by sampling effects due to the smaller number of markers for individual chromosomes. In addition, we used the Jaccard coefficient for AFLPs and the Dice coefficient for SSRs following the proposal by Link et al. (1995) for dominant and codominant marker systems, respectively. This may also have slightly decreased the correlation between AFLP- and SSR-derived GD values.

GD within inbred and DH lines

Inbred lines

In general, the dominant AFLP markers yielded lower GD_w values than the codominant SSR markers. As heterogeneity due to residual heterozygosity was the major cause of SSR variation within inbred lines, the lower GD_w values of AFLPs can be explained by the fact that heterogeneity cannot be detected using a dominant marker system like AFLPs. In addition, the

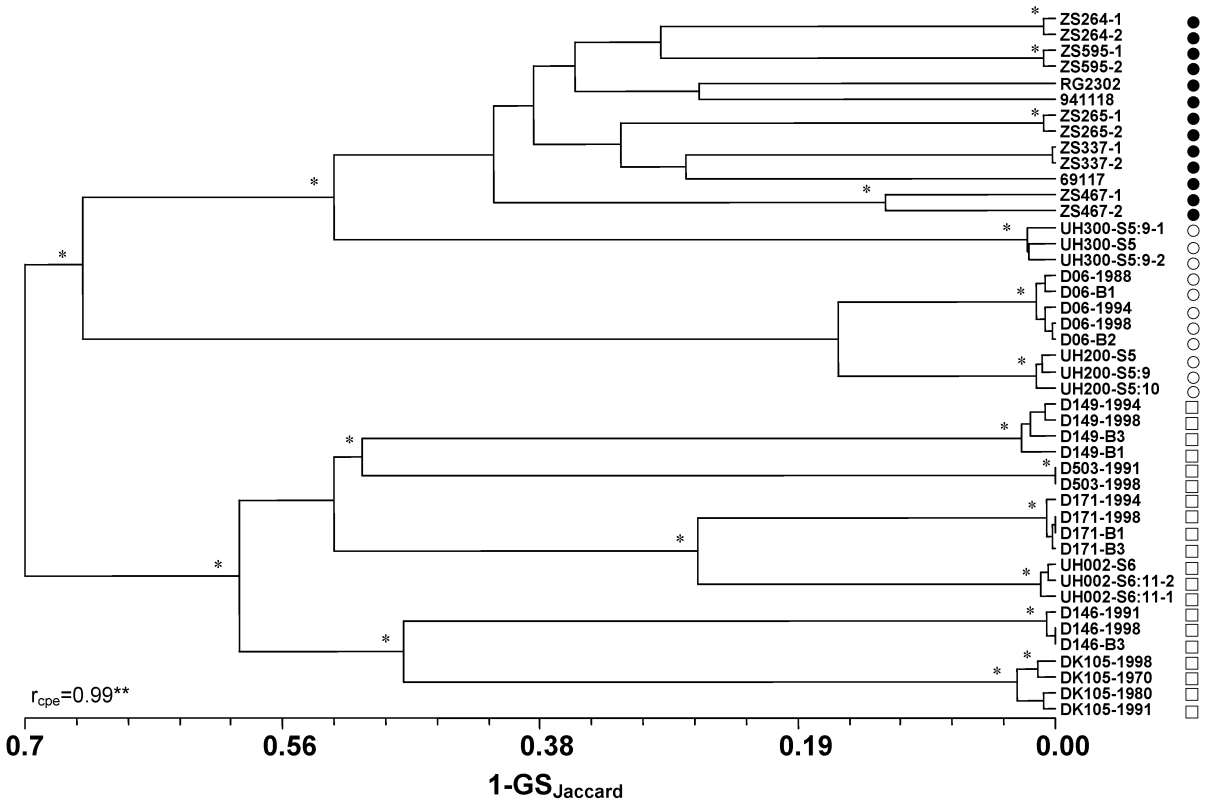


Figure 1. Associations among accessions of maize inbred lines revealed by UPGMA cluster analysis based on genetic distances calculated from AFLP data. Asterisks (*) at the forks indicate that the group to the right of the fork was found in at least 95% of 10000 bootstrap runs. DH lines and their parents are marked by filled circles (●). Flint and dent lines are marked with squares (□) and circles (●;○), respectively. DH lines were derived from F_1-hybrid 941118 generated by crossing lines s69117 and RG2302.

Table 3. Correlations between GD_{SSR} and GD_{AFLP} based on GDs of all 946 pairs of accessions or only GDs between 46 pairs of accessions from the same line (GD_w) for single chromosomes

	$r(GD_{SSR}, GD_{AFLP})$	
Chromosome	All GDs	GD_w
1	0.82**	0.01
2	0.91**	0.32*
3	0.84**	–0.02
4	0.87**	–0.19
5	0.84**	–0.16
6	0.81**	–0.03
7	0.89**	0.18
8	0.76**	0.02
9	0.83**	–0.08
10	0.60**	–0.10
All chromosomes	0.97**	0.07

generally lower degree of polymorphism of AFLPs compared to SSRs (Powell et al. 1996) may also contribute to the observed low GD_w values for AFLPs.

A variety of genetic and technical causes can contribute to the observed variation of GD_w estimates between accessions of the same line. First, the occurrence of point mutations is a possible cause. A mutation can result in a loss of an AFLP band, if it renders or disrupts the recognition site of the restriction enzyme or the selective bases of the primer. The natural mutation rate for higher eukaryotes was estimated to range from 10^{-9} to 10^{-7} per bp and generation (Drake et al. 1998). No data on mutation rates of AFLPs is yet available. However, if it is not significantly higher than the above mentioned values, mutations can explain only a minor fraction of the observed variation for GD_w values.

Second, lines in higher selfing generations of maintainance breeding are fixed for different alleles because the parental S_5 or S_6 lines were still heterozygous for a minor proportion of the genome.

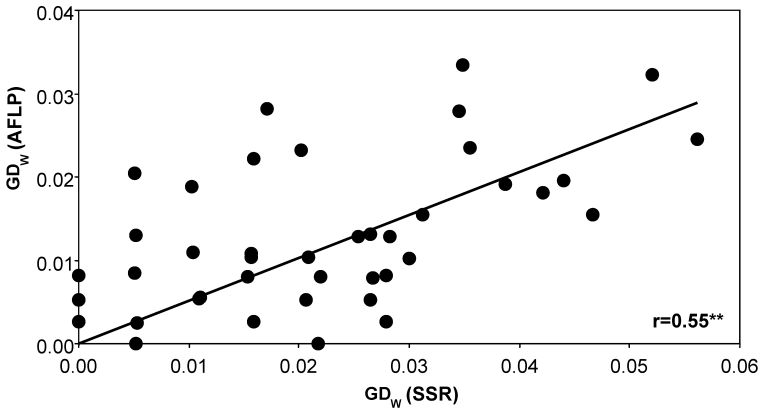


Figure 2. Plot of GD_w values revealed by 100 SSR markers and 20 AFLP primer combinations without outliers.

Third, contamination with foreign pollen could contribute to the observed variation within lines. However, this can be ruled out, because it should affect a large number of loci and, consequently, would be detectable as outliers. In addition, also our SSR results do not support the hypothesis of contamination with foreign pollen.

Fourth, heteroduplex bands can be a source for heterogeneity in SSR fingerprints (Hatcher et al. 1993). Heteroduplex bands emerge by annealing of two DNA fragments of unequal sequence or length and, therefore, tend to retard the migration of the particular band during electrophoresis (Hatcher et al. 1993; Nataraj et al. 1999). In AFLPs, no information on heteroduplex bands is available. Even though the occurrence of heteroduplexes in AFLPs cannot be ruled out as a source of variation in GD estimates, heteroduplex AFLP fragments were never encountered when the sequences derived from both strands of single AFLP markers were compared (J. Rouppe van der Voort and J. Peleman, pers. comm.).

In contrast, genotyping samples from bulked individuals may slightly decrease the variation within lines.

DH lines

DH lines were included in our study to distinguish between genetic and technical reasons for variation of GD within accessions of the same line. Since DHs are homogeneous and genetically uniform, differences among samples are only attributable to technical causes. While SSRs yielded identical fingerprints, replications of DH lines were not scored identical with AFLPs. As the seed samples of the two accessions fingerprinted in our study were drawn out of the same seed lot, they should be scored as identical. Therefore, segregation and bulking effects can be ruled out as reasons for the observed variation of AFLPs within DH lines.

Technical variation due to poor DNA quality, incomplete digestion of DNA, inconsistent amplification, or scoring problems of the applied software are the most probable reasons for the observed variation between identical samples of DH lines. In addition, heterogeneity within DH lines was observed by Murigneux et al. (1993). However, this is very unlikely, as our SSR results do not support this hypothesis.

Further investigations revealed that the DNA of accession ZS467-2 was incompletely digested, which explains the extraordinarily high variation between the two accessions of this DH line. This indicates that even for a highly reproducible marker technique such as AFLPs, routine analyses could lead to incorrect results in the case of a suspected EDV. Therefore, we recommend replication of the lab assays to minimize the experimental error. Given the high value of litigation involved in EDV claims, additional costs for replicated lab assays are well justified.

Locations of variation

Matches in the locations of variation detected with AFLPs and SSRs could be caused by the fact that certain genomic regions were still segregating by the time the accessions were separated. In addition, it cannot be ruled out that these matches are attributable to chance. However, when the BIN positions of markers contributing to the observed variation within lines were compared, only a low coincidence between the

locations of variation for AFLPs and SSRs was detected. For example, chromosome 9 revealed the highest number of SSRs displaying variation between accessions of the same line but the lowest for AFLPs. This observed discrepancy is presumably attributable to the different methods by which dominant AFLPs and codominant SSRs generate polymorphisms. In addition, the number of markers displaying variation within maize lines was too small in order to assign the observed variation unambiguously to certain genomic regions. Therefore, inferences about genomic regions showing variation in GD_w estimates must be considered with caution.

Use of AFLPs and SSRs for identification of EDVs

The rationale underlying the use of molecular markers in testing for EDVs is to take the GD estimated from marker data as an indicator for the true GD between two genotypes across the entire genome. If the estimated GD exceeds a certain threshold, the two genotypes are considered as independently derived, whereas otherwise this is taken as evidence for a putative EDV. Obviously, any variation between different accessions of the same inbred or DH line attributable to genetic and/or technical reasons is not only reflected by the GD_w but will also affect the GD between accessions of different lines. High levels of variation within lines have as consequence a decreased resolution to distinguish related lines. In general, this type of variation will inflate the Type I error in testing for essential derivation (H_0: One of the two lines is an EDV of the other) and reduce the Type II error. If the range of GD values between accessions of the same line is large, it might be possible that a true EDV could be judged as independently derived or a truly independent variety as essentially derived just by genotyping different accessions of each variety (Heckenberger et al. 2002). Subsequently, we compare RFLP, RAPD, AFLP, and SSR markers under this aspect and other criteria relevant in investigations of EDVs.

First, the reproducibility of molecular marker data using different sources of DNA extracted from the same accession or across laboratories must be high. This criteria excludes RAPDs from the list of possible marker systems because they were found to be less reproducible than AFLPs or SSRs (Bagley et al. 2001; Jones et al. 1997).

Second, the applied marker system should reveal a high degree of polymorphism. In the ideal case, if unrelated genotypes have no marker bands in common (GD = 1), identical bands in related individuals are exclusively attributable to 'identity by descent' and, thus, directly reflect the degree of relatedness. However, with GD < 1 for unrelated genotypes, some marker bands are 'identical in state' and this must be taken into account when defining EDV thresholds. In our study, the mean GD for unrelated dent lines was significantly different between AFLPs and SSRs suggesting that the breeders must agree on different EDV thresholds for each marker system.

Third, the applied marker system should warrant at low costs a uniform and dense coverage of the entire genome to obtain unbiased GD estimates with small standard errors. In principle, this requirement can be met by AFLPs, RFLPs, and SSRs, but AFLPs offer the advantage that a large number of markers can be produced per primer combination. In comparison with AFLPs, RFLPs and SSRs have a lower MI, but their information content is higher due to the codominant inheritance and the higher degree of polymorphism. In addition, detailed marker information, including primer sequences and map position, are publicly available for RFLPs and SSRs in maize (see MaizeDB http://nucleus.agron.missouri.edu/index.html, confirmed 16 April 03). Owing to limitations in the automation and standardization of RFLPs, they represent no longer a competitive alternative to SSRs and AFLPs for EDV analyses.

AFLP markers can be produced at lower costs per data point than SSRs, because one primer combination produces a large number of bands. Despite this economic advantage of AFLPs, the French association of maize breeders (SEPROMA) recently recommended a set of uniformly distributed SSR markers for EDV identification. SSRs provide a higher degree of transparency for legal issues than AFLPs due to their codominant inheritance, their known map positions, and their public availability. However, since the set of SSRs is known to all breeders, it is, in principle, possible to use this information to select for genetic diversity at some SSR markers to avoid an EDV, while maintaining a high degree of relatedness in other genomic regions. Complementary use of AFLP markers would prevent this situation, because selection for variants at specific marker loci would be more difficult for AFLPs than for a fixed set of SSRs. In addition, the redundancy in the genotypic information caused by clustered AFLPs could be omitted by applying map based genetic distances (Dillmann et al. 1997).

Taking into account all criteria, both AFLPs and SSRs are suitable marker systems for EDV identification. In order to counterbalance advantages and disadvantages of each marker system, AFLPs and SSRs could be used in a complementary way to unambiguously distinguish EDVs and independent varieties. We further conclude that the stability of marker data across different generations of multiplication or accessions maintained by different breeders is primarily a function of marker reproducibility and residual heterozygosity or heterogeneity. Therefore, with regard to the use of DNA markers for resolving EDV issues, it is important to reduce residual heterozygosity before applying for plant variety protection. This can be achieved by further selfing and/or pre-screening of lines with molecular markers for homogeneity or by production of DH lines.

Acknowledgements

The present study was supported by a grant from the European Union, Grant No. QLK-CT-1999-01499 (MMEDV) (for further information, see http://www.niab.com/bbp/mmedv/, confirmed 16 April 03). In addition, the authors wish to thank Semundo (Hadmersleben, Germany) for supplying the DH lines.

References

Anscombe F.J. and Tukey J.W. 1963. The examination and analysis of residuals. Technometrics 5: 141–160.

ASSINSEL 2000. DUS testing: Phenotype vs. Genotype. ASSINSEL, Position Paper adopted at the Rome Congress in May 2000.

Bagley M.J., Anderson S.L. and May B. 2001. Choice of methodology for assessing genetic impacts of environmental stressors: polymorphism and reproducibility of RAPD and AFLP fingerprints. Ecotoxicology 10: 239–244.

Bar-Hen A. and Charcosset A. 1994. Relationship between molecular and morphological distances in a maize inbred lines collection. Application for breeder's rights protection. Biometrics in Plant Breeding: Applications of Molecular Markers. In: Proceedings of the Ninth Meeting of the EUCARPIA section Biometrics in Plant Breeding. pp. 57–66.

Dice L.R. 1945. Measures of the amount of ecologic association between species. Ecology 26: 297–302.

Dillmann C., Bar-Hen A., Guerin D., Charcosset A. and Murigneux A. 1997. Comparison of RFLP and morphological distances between maize (*Zea mays* L.) inbred lines – consequences for germplasm protection purposes. Theor. Appl. Genet. 95: 92–102.

Drake J.W., Charlesworth B., Charlesworth D. and Crow J.F. 1998. Rates of spontaneous mutation. Genetics 148: 1667–1686.

Efron B. 1979. Bootstrap methods: another look at the jackknife. Ann. Stat. 7: 1–26.

Hansen M., Kraft T., Christiansson M. and Nilsson N.O. 1999. Evaluation of AFLP in *Beta*. Theor. Appl. Genet. 98: 845–852.

Hatcher S.L., Lambert Q.T., Raymond L.T. and Carlson J.R. 1993. Heteroduplex formation: a potential source of errors from PCR products. Prenatal Diagnosis 13: 171–177.

Heckenberger M., Bohn M., Ziegle J.S., Joe L.K., Hauser J.D., Hutton M. et al. 2002. Variation of DNA fingerprints among accessions within maize inbred lines and implications for identification of essentially derived varieties. I. Genetic and technical sources of variation in SSR data. Mol. Breed. 10: 181–191.

Jaccard P. 1908. Nouvelles recherches sur la distribution florale. Bull. Soc. Vaud. Sci. Nat. 44: 223–270.

Johnson N.L., Kotz S. and Kemp A.W. 1992. Univariate Discrete Distributions. J. Wiley and Sons, Inc., New York.

Jones C.J., Edwards K.J., Castaglione S., Winfield M.O., Sala F., Vandewiel C. et al. 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Mol. Breed. 3: 381–390.

Knaak C., Förster J. and Jäger-Gussen M. 1996. 'Abgeleitete' Sorten aus praktischer Sicht. Bericht über die 47. Arbeitstagung 1996 der Saatzuchtleiter im Rahmen der 'Vereinigung österreichischer Pflanzenzüchter' BAL Gumpenstein. pp. 167–172.

Link W., Dixkens C., Singh M., Schwall M. and Melchinger A.E. 1995. Genetic diversity in European and mediterranean faba bean germ plasm revealed by RAPD markers. Theor. Appl. Genet. 90: 27–32.

Lübberstedt T., Melchinger A.E., Dussle C., Vuylsteke M. and Kuiper M. 2000. Relationships 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.

Lübberstedt T., Vuylsteke M., Dussle C., Kuiper M. and Melchinger A.E. 1999. Untersuchungen zur genetischen Diversität zwischen europäischen Maisinzuchtlinien mit AFLPs, RFLPs, RAPDs, und SSRs – ein Vergleich verschiedener Markersysteme. Bericht über die 49. Arbeitstagung 1998 der Vereinigung österreichischer Pflanzenzüchter, BAL Gumpenstein. pp. 45–49.

Mantel N.A. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. 27: 209–220.

Melchinger A.E. 1999. Genetic Diversity and Heterosis. In: Coors J.G. and Pandery S. (eds), The Genetics and Exploitation of Heterosis in Crops. Crop Science Society of America, Inc., Madison, USA, pp. 99–118.

Murigneux A., Barloy D., Leroy P. and Beckert M. 1993. Molecular and morphological evaluation of doubled haploid lines in maize. 1. Homogeneity within DH lines. Theor. Appl. Genet. 86: 837–842.

Nataraj A.J., Olivos-Glander I., Kusukawa N. and Highsmith W.E. 1999. Single-strand conformation polymorphism and heteroduplex analysis for gel-based mutation detection. Electrophoresis 20: 1177–1185.

Nei M., Tajima F. and Tateno Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. J. Mol. Evol. 19: 153–170.

Peleman J., van Wijk R., van Oeveren J. and van Schaik R. 2000. Linkage map integration: an integrated genetic map of *Zea mays* L. Poster P472. Plant & Animal Genome Conference VIII.

Powell W., Morgante M., Andre C., Hanafey M., Vogel J., Tingey S. et al. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol. Breed. 2: 225–238.

Rohlf F.J. 1989. NTSYS-pc Numerical Taxonomy and Multivariate Analysis System. Exeter Publishing Co, Ltd., Setauket, NY, USA.

Roldan-Ruiz I., Calsyn E., Gilliland T.J., Coll R., van Eijk M.J.T. and De Loose M. 2000. Estimating genetic conformity between related ryegrass (*Lolium*) varieties. 2. AFLP characterization. Mol. Breed. 6: 593–602.

SAS Institute 1988. SAS/STAT User's guide, Release 6.03 edn. SAS, Cary.

Smith J.S.C., Smith O.S., Bowen S.L., Tenborg R.A. and Wall S.J. 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 RFLPs. Maydica 36: 213–226.

Vos P., Hogers R., Bleeker M., Reijans M., Lee T.v.de., Hornes M. et al. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23: 4407–4414.

Yap I.V., Nelson R.J. and WinBoot 1996. A program for performing bootstrap analysis of binary data to determine confidence limits of UPGMA-based dendrograms. International Rice Research Institute, Manila, Philippines.

4. Identification of essentially derived varieties (EDVs) derived from biparental crosses of homozygous lines. I. SSR data from maize inbreds

M. Heckenberger, M. Bohn, and A. E. Melchinger

Abstract

Genetic distances (GDs) based on molecular markers such as simple sequence repeats (SSRs) have been proposed as an appropriate tool to assess the genetic conformity between putative essentially derived varieties (EDVs) and their initial varieties (IVs). However, for maize and other crops no consensus has been reached regarding GD thresholds for identification of EDVs, because reliable benchmark data are lacking. The objectives of this study were to (1) estimate the variation in the parental contribution (*p*) to the genome of homozygous progeny lines derived in recycling breeding programs, (2) investigate the power of SSR-based GD estimates for discriminating between progeny lines derived from F_2, BC_1, and BC_2 populations, (3) exemplify the theoretical and simulated results of a companion study with experimental data, and (4) draw conclusions with regard to various EDV thresholds suggested hitherto. A total of 220 European and U.S. maize inbred lines comprising 163 triplets were genotyped with 100 uniformly distributed SSRs. A triplet consisted of one F_2-, or BC_1-progeny line and both parental lines. SSR-based estimates of *p* varied from 0.25 to 0.74 for F_2-derived lines with a mean (0.49) close to the expectation (0.50) and ranged from 0.51 to 0.80 for BC_1-derived lines with a mean (0.66) significantly smaller than the expectation (0.75). Relative to the variation in *p*, the GD between progeny lines and parents was less influenced by the variation in the GD between the parents, particularly for BC_1-derived lines. Suggested GD thresholds T for EDVs using a fixed GD yielded considerably different values for Type I (α) and Type II ($1-\beta$) errors among different gene pools and material groups. Therefore, we recommend germplasm specific thresholds with fixed α or $\alpha=1-\beta$.

Introduction

Legal regulations for plant variety protection (PVP) should secure the reward for past breeding efforts but also sustain future breeding progress. Registered plant varieties need to be protected against plagiarism and misuse on the one hand, but protected germplasm should be accessible for the development of new varieties on the other hand. The latter was warranted by the concept of "breeder's exemption" or "breeder's privilege" in the original convention of the Union for the Protection of New Varieties of Plants (UPOV, 1978).

The advent of new methods such as genetic engineering and marker-assisted backcrossing, however, has provided the basis to undermine the breeder's exemption in its original intention. These tools make it possible to add a few new genes to a protected variety or to select deliberately for lines that are very similar to one of their parents and apply for PVP for this "new" variety. Therefore, the investments made in breeding the original variety can be exploited by the breeder of the plagiarized variety without indemnification for the breeder of the original variety.

The concept of essentially derived varieties (EDVs) was implemented into the revised UPOV convention (UPOV, 1991) and several national PVP acts to cope with this new situation. Accordingly, a variety is deemed to be essentially derived from an initial variety (IV), if it (i) was predominantly derived from the IV, (ii) is clearly distinguishable from the IV, and (iii) is genetically conform to the IV. However, breeding companies have not agreed on specific breeding procedures that are considered to yield independently derived varieties (IDVs) or EDVs (*e.g.*, the number of acceptable backcross generations to a protected variety). In addition, no official guidelines or appropriate methods have been fixed to assess the genetic conformity between IVs and potential EDVs. Hence, crop-specific thresholds for the discrimination between EDVs and IDVs have not yet been defined.

In principle, the coefficient of parentage (*f*) introduced by Malécot (1948) could serve for identification of EDVs, because it reflects the degree of relatedness between two genotypes on the basis of their pedigree. In the case of a suspected EDV, however, pedigree

data are usually not available for the breeder of the IV. In addition, *f* is an indirect measure of genetic similarity based on several simplifying assumptions such as equal parental genome contributions and absence of selection, mutation, or drift (Messmer et al., 1992).

Molecular markers such as simple-sequence repeats (SSRs) or amplified fragment length polymorphisms (AFLPs) allow to determine the parental origin of the chromosomal segments in a progeny. Therefore, genetic distances (GDs) based on molecular markers were proposed as an appropriate tool to determine the genetic conformity between an IV and putative EDVs and, consequently, to distinguish between EDVs and IDVs (ASSINSEL, 1999; International Seed Federation, 2002). In maize, GDs between lines based on AFLP and SSR data were tightly correlated with each other and with *f* estimates (Lübberstedt et al., 2000; Smith et al., 1997), suggesting that the degree of relatedness of two genotypes can be inferred from their GD. However, distributions of GDs for F_2- and BC_1- derived progenies showed a substantial overlap (Bernardo et al., 1997).

In a companion paper, we proposed a conceptual framework, based on principles of statistical test theory, for identification of EDVs with molecular markers (Bohn et al., 2003). Accordingly, for a progeny line derived from bi-parental crosses, the GD to each parent depends on the GD between the two parents and *p*, the parental genome contribution transmitted to the progeny. Experimental estimates of *p* for F_2- and BC_1-derived progenies were reported by Bernardo et al. (1997; 2000). Moreover, formulas for the variance of *p* for both types of progeny were derived by Wang and Bernardo (2000). None the less, further experimental data are required to verify the approach of Bohn et al. (2004) and quantify the influence of the above mentioned factors with regard to consequences for potential EDV thresholds.

In this study, we investigated a large number of triplets in maize, each consisting of homozygous progeny lines derived from F_2, BC_1, or BC_2 populations and their parental inbreds. Our objectives were to (1) estimate the variation in the parental contribution to the genome of the progeny, (2) investigate the power of SSR-based GD estimates for discriminating between progenies derived from F_2, BC_1, and BC_2 populations, (3) exemplify the theoretical and simulated results of Bohn et al. (2004) with experimental data, and (4) draw conclusions with regard to various EDV thresholds suggested in the literature.

Materials and Methods

Plant Materials

A total of 220 elite maize inbred lines were analyzed comprising 89 European flint, 74 European dent, 14 U.S. dent, and 43 introgression lines. These lines originated from the maize breeding programs at the University of Hohenheim (Stuttgart, Germany), Iowa State University (Ames, USA), and three commercial breeding companies in Germany. The 220 lines comprised 163 triplets. A triplet consisted of one progeny and both parental lines. The materials consisted of 118 intra-pool triplets of European dent or flint lines and 45 inter-pool triplets, each consisting of one European and one U.S. line with an introgression line as progeny. Altogether, 83% of the progenies were derived from F_2 populations and 17% were derived from BC_1 or BC_2 populations (Table 1). Detailed information on all 163 triplets and the 220 maize inbreds included in this study is available in Tables A and B in the appendix of this thesis.

Molecular Analyses

All lines were genotyped with a set of 100 SSR markers uniformly covering the entire maize genome as described in detail by Heckenberger et al. (2002). Briefly, DNA samples were analyzed using an ABI Prism™ 377 DNA Sequencer with 96 lane polyacrylamid gels. Internal fragment size standards were used in each lane to increase accuracy of DNA fragment size determination. The size of each DNA fragment was determined automatically by using the GeneScan® software and assigned to specific alleles by the Genotyper® software. The 100 SSRs were selected on the basis of reliable single-locus amplification, absence of null alleles, high degree of polymorphism, and high reproducibility of the bands. Seventy of the 100 SSRs contained di-nucleotide repeat motifs, whereas the other 30 markers consisted of tri- to octa-nucleotide repeats. SSR analyses were performed on a commercial basis by Celera (1756 Picasso Avenue, Davis CA 95616, USA). Non-parental alleles were defined as alleles present in the progeny line, but absent in each of the parents.

Statistical Analyses

Malécot's (1948) coancestry coefficient (*f*) was calculated between all pairwise line combinations. Genetic distances (GDs) between lines were estimated using Rogers' distance (Rogers, 1972). In the case of missing values in one of the two inbreds compared, the corresponding alleles of the other accession were not used for GD calculation. Standard errors (SEs) for GDs were estimated using the jackknife procedure (Efron, 1979) with resampling over primer pairs (Tivang et al., 1994). Coefficients of correlation between GD_{SSR} and *f* were calculated using simple correlation coefficients (Snedecor and Cochran, 1980). In addition, the linear relationship between *f* and GD was tested with a lack-of-fit test. Calculation of GDs were performed with the PLABSIM software (Frisch et al., 2000). All other statistical calculations were performed with the R software package (Ihaka and Gentleman, 1996).

Suppose progeny line O is derived from a biparental cross (*e.g.,* F_2, BC_1, or BC_2 generation) between the homozygous parent P1 and P2 and the GDs between P1 and P2 or O are denoted by $GD_{(P1,P2)}$ and $GD_{(P1,O)}$, respectively. When O was an F_2-derived homozygous progeny line, P1 was the first parent listed in the pedigree record of O. When O was a BC_1-derived inbred, P1 was the recurrent parent, whereas P2 was the donor parent. If GD is determined by a large number of polymorphic markers with uniform coverage of the entire genome, we obtain the following equation:

$$GD_{(P1,O)} = (1-p)GD_{(P1,P2)}, \qquad [1]$$

where *p* denotes the proportion of the genome transmitted from P1 to O.

Solving Eq. [1] for *p* yields

$$p = 1 - \frac{GD_{(P1,O)}}{GD_{(P1,P2)}}, \qquad [2]$$

which can be used for estimating *p*. Similar formulas were given by Bernardo et al. (1997, 2000) on the basis of the number of common bands between P1 and O or the simple matching coefficient (Sneath and Sokal, 1973). Since the latter is based on single alleles without weighting of multiple bands within a marker locus, we chose the Rogers' distance for this study.

In the absence of selection, *p* is a random variable with distribution properties depending on (a) the degree of relatedness between P1 and O and (b) the number and length of the chromosomes (Wang and Bernardo, 2000). If P1 and P2 are unrelated ($f_{(P1,P2)}$=0), then the expected value μ_p of *p* corresponds to the coancestry $f_{(P1,O)}$ and, thus, μ_p=0.500, 0.750, and 0.875 for F_2-, BC_1-, or BC_2-derived progeny lines of P1, respectively.

Formulas for the variance σ_p^2 of F_2- or BC_1-derived progeny lines were given by Wang and Bernardo (2000). In addition, numerical values for maize were obtained for F_2-, BC_1-, or BC_2-derived progeny lines from stochastic simulations by Bohn et al. (2004). The latter were based on a genetic model allowing for genetic drift but neither selection nor mutation. Hence, empirical and simulated frequency distributions of *p* values were compared with a Kolmogorov-Smirnov test (Lehmann, 1986) to check for significant deviations caused by selection or mutations. Equality of variances of empirical and simulated frequency distributions of *p* was evaluated with Levene's test (Levene, 1960).

If progeny lines are derived from a large number of bi-parental crosses with different pairs of parents P1 and P2 representative for a germplasm pool, then $GD_{(P1,P2)}$ can be regarded as a random variable with mean $\mu_{GD_{(P1,P2)}}$ and variance $\sigma^2_{GD_{(P1,P2)}}$. Since the value of *p* for a specific progeny is completely unrelated to the GD of its parent lines, $GD_{(P1,P2)}$ and *p* are stochastically independent. Thus, we obtain from Eq. [1] the following equations (Bohn et al., 2004):

$$\mu_{GD_{(P1,O)}} = \mu_{GD_{(P1,P2)}} - \mu_{GD_{(P1,P2)}}\mu_p \qquad [3]$$

$$\sigma^2_{GD_{(P1,O)}} = (1-\mu_p)^2\sigma^2_{GD_{(P1,P2)}} + \mu^2_{GD_{(P1,P2)}}\sigma^2_p + \sigma^2_{GD_{(P1,P2)}}\sigma^2_p, \qquad [4]$$

where $\mu_{GD_{(P1,O)}}$ and $\sigma^2_{GD_{(P1,O)}}$ are the mean and variance of $GD_{(P1,O)}$, respectively, for a given relationship between O and P1. By inserting experimental estimates for $\sigma^2_{GD_{(P1,P2)}}$ and estimates for μ_p and σ_p^2 determined (a) either from computer simulations (Bohn et al., 2004) or (b) the formulas given by Wang and Bernardo, (2000), we were able to calculate predicted values for $\sigma^2_{GD_{(P1,O)}}$ and compare them with estimated values for F_2- or BC_1-derived progeny lines from unrelated parents. In addition, Eq.[4] permits to compare the relative

influence of σ_p^2 and $\sigma_{GD_{(P1,P2)}}^2$ on the variance of $GD_{(P1,O)}$ for F_2- or BC_1-derived progeny lines, which is of importance for the question of EDV thresholds. In addition, simulated $GD_{(P1,O)}$ values were calculated with Eq. [1] for each material group on the basis of simulated *p* values and $\hat{\mu}_{GD_{(P1,P2)}}$ and $\sigma_{GD_{(P1,P2)}}^2$ for observed $GD_{(P1,P2)}$ values of unrelated lines.

Threshold Scenarios

To increase the sample size, not only GD values obtained within triplets were used for evaluation of potential thresholds (T), but all GD values of the dataset with corresponding *f* values of 0.500, 0.750, and 0.875 for F_2-, BC_1-, or BC_2-derived progeny lines. The frequency distributions of empirical $GD_{(P1,O)}$ values for F_2-, BC_1-, or BC_2-derived progeny lines were approximated by beta distributions (Johnson et al., 1995) with parameters chosen such that the mean and variance of the original distribution were conserved. Based on these distributions, we calculated Type I (α) and Type II (1-β) errors for various EDV thresholds T and various types of populations. Here, α corresponds to the probability that a true IDV will be wrongly judged as EDV and 1-β corresponds to the probability that a true EDV will not be recognized as such and judged as IDV (Fig. 1). First, we considered the situation that an F_2-derived progeny will be regarded as IDV, but a BC_1-derived progeny as EDV. Second, we assumed that a BC_1-derived progeny will be regarded as IDV, but a BC_2-derived progeny as EDV.

SSR- or RFLP-based GD values of 0.25, 0.20, 0.15, and 0.10 were suggested as possible EDV thresholds T by ASTA (Smith and Smith, 1989), ASSINSEL (2000), SEPROMA (Leipert, 2003, personal communication) and Troyer and Rocheford (2002), respectively. For all thresholds, the corresponding α and 1-β values were calculated for homozygous progeny lines derived from F_2, BC_1, and BC_2 populations. In addition, other thresholds T with fixed α=0.05 ($T_{0.05}$) or α=1-β ($T_{\alpha=1-\beta}$) were tested.

Results

Genetic Variation for SSRs

A total of 1099 SSR alleles were observed with the 100 SSRs on the set of 220 inbred lines. The number of alleles per marker varied from 3 to 26. PIC values ranged from 0.10 to 0.88 with a mean of 0.71. Only 3.7% of all marker data points were missing due to amplification failure or null alleles. Correlations between GD and *f* were highly significant ($P<0.01$) for all three material groups and highest for dent lines (r=-0.90**), intermediate for flint lines (r=-0.75**) and lowest for introgression lines (r=-0.58**). In addition, we observed a linear relationship between *f* and GD for all three material groups. A detailed description of the genetic diversity of the germplasm is given elsewhere.

Parental Contributions (*p*) for F_2- and BC_1-derived Progenies

The three material groups did not differ from each other in their means $\hat{\mu}_p$ for both the F_2- and BC_1- derived progenies. Hence, the data from all three groups were pooled for further analyses. For F_2-derived progenies, SSR-based estimates of *p* ranged from 0.25 to 0.74 with $\hat{\mu}_p$=0.49 (Fig. 2), close to the expectation of 0.50. Variances for observed and simulated values of *p* (σ_p^2) did not differ significantly ($P<0.05$) (Table 2). Frequency distributions for observed and simulated estimates of *p* were significantly different ($P<0.05$) from each other due to a higher kurtosis of the former.

SSR-based estimates of *p* for BC_1-derived progenies varied from 0.51 to 0.80 with a mean $\hat{\mu}_p$=0.66, which was significantly smaller than the expectation of 0.75 (Fig. 2). Variances for observed and simulated values of *p* (σ_p^2) were not significantly ($P<0.05$) different from each other (Table 2). Frequency distributions for observed and simulated estimates of *p* showed significant differences ($P<0.01$) due to the shift to smaller values, the lower skewness and the higher kurtosis for the distribution of observed *p* values.

Genetic Distances among Unrelated Parental Inbred Lines

GDs among unrelated ($f_{(P1,P2)}$=0) flint lines ranged from 0.23 to 0.79 with $\hat{\mu}_{GD_{(P1,P2)}}$=0.58 (Fig. 3). GDs for unrelated dent lines varied from 0.25 to 0.85 with a significantly (P<0.01) larger mean $\hat{\mu}_{GD_{(P1,P2)}}$=0.61. Unrelated parents of introgression lines, consisting of pairs of European and U.S. maize lines, had by far the largest range from 0.22 to 0.93 and also a significantly (P<0.01) higher mean $\hat{\mu}_{GD_{(P1,P2)}}$=0.74 than the intra-pool pairs of the other two material groups.

Subdivision of the Variance of $GD_{(P1,O)}$ for F_2- and BC_1-derived Progenies

Observed values of $\sigma^2_{GD_{(P1,O)}}$ obtained directly from experimental data were in close agreement with the predicted $\sigma^2_{GD_{(P1,O)}}$ values calculated with Eq.[4] on the basis of simulated values of μ_p and σ^2_p as well as experimental estimates of $\hat{\mu}_{GD_{(P1,P2)}}$ and $\sigma^2_{GD_{(P1,P2)}}$. Further analysis revealed that for F_2-derived progenies 65% of $\sigma^2_{GD_{(P1,O)}}$ could be explained by σ^2_p and 34% by $\sigma^2_{GD_{(P1,P2)}}$. For BC_1-derived progenies, 94% of $\sigma^2_{GD_{(P1,O)}}$ were explained by σ^2_p and only 5% by $\sigma^2_{GD_{(P1,P2)}}$. The contribution of the product $\sigma^2_p\sigma^2_{GD_{(P1,P2)}}$ to $\sigma^2_{GD_{(P1,O)}}$ was less than 1% for both F_2- and BC_1-derived progeny lines (Table 2).

Evaluation of EDV-Threshold Scenarios

Observed frequency distributions of GD values for F_2-, BC_1-, and BC_2-derived progenies fitted well the approximated beta distributions for flint and dent lines, but only moderately for introgression lines (Fig. 4). For all three material groups, considerable overlaps between the frequency distributions of GDs for F_2- *vs.* BC_1- as well as for BC_1- *vs.* BC_2-derived progenies were observed. Within each generation, $\hat{\mu}_{GD_{(P1,O)}}$ was significantly higher (P<0.05) for the dent lines than for the flint lines. In addition, $\hat{\mu}_{GD_{(P1,O)}}$ for the introgression lines was always significantly higher (P<0.01) than $\hat{\mu}_{GD_{(P1,O)}}$ for the flint and dent

lines. Estimates of $\sigma^2_{GD_{(P1,O)}}$ for the same generation were not significantly different (P<0.01) between flint and dent lines but significantly (P<0.01) larger for introgression lines.

Given α=0.05 for F_2-derived lines, the power β to classify a BC_1-derived progeny line as EDV amounted to 77%, 63%, and 15% for the particular thresholds determined for flint, dent, and introgression lines, respectively (Table 3.). Corresponding values of β for BC_2-derived lines, assuming α=0.05 for BC_1-derived lines were smaller for flint and dent lines, but larger for introgression lines. The power β for thresholds determined for α=1-β to classify BC_1- or BC_2-derived progenies as EDVs increased considerably compared to the values for α=0.05. This increase in the power β, however, is associated with higher values for α. Therefore, this leads to a considerably higher frequency of F_2- or BC_1-derived progenies incorrectly classified as EDVs.

For T=0.25, 0.20, or 0.15, the corresponding α levels for F_2-derived lines varied between α=0.18 and α=0.00 (Table 3). Corresponding β values ranged between 7% and 92%. For T=0.15 and T=0.10, the power β to detect a BC_2-derived line as EDV varied from 10% to 99% with corresponding α values for BC_1-derived lines ranging from 0.02 to 0.07. For each T substantial differences for α and β between flint, dent, and introgression lines were observed.

For α=0.05 and α=1-β, T values obtained from simulated data were lower than from observed data with the exception of α=0.05 for introgression lines (Table 3.). For all these scenarios, the power β to classify BC_1- or BC_2-derived progeny lines as EDVs was similar between thresholds based on observed and simulated values of $GD_{(P1,P2)}$ for both flint and dent lines. For introgression lines, however, β was substantially higher for T values based on simulated data than those based on observed data. Considerable differences existed also between observed and simulated data regarding values of α and β for T=0.25, 0.20, 0.15 and 0.10.

Discussion

Our study was initiated by commercial breeding companies to derive EDV thresholds in maize based on scientifically reliable criteria, as requested by UPOV and ASSINSEL. Representative germplasm for each material group was taken from public and private breeding programs. SSRs were chosen as a suitable marker system due to their known map positions, high degree of polymorphism and suitability for automated high-throughput analyses. Therefore, our results are relevant for the definition of EDV thresholds and provide a general overview on putative essential derivation scenarios in European maize germplasm and the power of SSRs for identification of EDVs.

Use of SSR-Based GDs for Identification of EDVs

The rationale for using SSR-based GD estimates for identification of EDVs is their close relationship to *f*. Therefore, they can be used to uncover close pedigree relationships between pairs of inbred lines. Correlations between GDs and *f* calculated across the entire data set (r=0.77) and separately for each material group were similar or higher than reported in previous studies with maize (Lübberstedt et al., 1999; Pejic et al., 1998). This reflects the broad basis of germplasm in this study ranging from unrelated to closely related combinations of lines. Moreover, the linear relationship of GD and *f* corroborates that GDs based on SSRs faithfully reflect the genetic diversity of the germplasm. In spite of the observed high correlations, considerable variation was observed for GD values obtained for the same *f* values and, thus, overlaps in the frequency distributions of GDs occurred for *f*=0.50, 0.75, and 0.88. Therefore, F_2-, BC_1-, and BC_2-derived progenies could not be distinguished unambiguously by their $GD_{(P1,O)}$.

Factors Influencing $GD_{(P1,O)}$

According to Eq. [1], $GD_{(P1,O)}$ is influenced by two factors: $GD_{(P1,P2)}$ and *p*. Assuming the ideal case that unrelated lines ($f_{(P1,P2)} = 0$) show a GD of 1.0, $GD_{(P1,O)}$ yields an estimate of 1-*p,* wich theoretically results in the highest discrimination ability between different types of progeny. However, even for this most favorable case, considerable overlaps

between the frequency distributions of F_2- and BC_1-derived or between BC_1- and BC_2-derived lines were found in simulations (Fig. 1).

Means and variances for distributions of observed *p* values for F_2-derived progenies were to a large extent identical with the distribution of simulated *p* values. However, the observed $\hat{\mu}_p$ for BC_1-derived progenies was substantially lower than the expectation (Table 2). This shift towards the distribution of F_2-derived progenies is very likely attributable to the selection of the most vigorous BC_1 plants in the development of improved progeny lines. Due to the phenomenon of heterosis, such BC_1 plants are more heterozygous and consequently have a higher proportion of donor genome than the average. Obviously this selection for more heterozygous plants would result in an increased overlap in the frequency distributions of GDs between F_2- and BC_1-derived or between BC_1- and BC_2-derived lines, compared to the simulated data shown in Figure 1.

Further comparison of the above-mentioned ideal case with authentic data revealed that $GD_{(P1,P2)}$ between unrelated lines was lower than 1.0 and showed a considerable variance $\sigma^2_{GD_{(P1,P2)}}$. This leads to condensed and more flat frequency distributions for $GD_{(P1,O)}$ values of F_2-, BC_1-, and BC_2-derived progenies and, therefore, to a further increase of the overlaps. The magnitude of the overlaps is mainly caused by the parameters $\hat{\mu}_{GD_{(P1,P2)}}$ and $\sigma^2_{GD_{(P1,P2)}}$ of unrelated lines. Due to different levels of genetic diversity among breeding germplasm of crops, $\hat{\mu}_{GD_{(P1,P2)}}$ and $\sigma^2_{GD_{(P1,P2)}}$ vary considerably among different crop species. For example, the GDs between unrelated barley (Melchinger et al., 1994) or tomato cultivars (Grandillo et al., 1999) were substantially lower than observed in maize (Messmer et al., 1993). This underlines the necessity of crop-specific thresholds T for the discrimination of EDVs and IDVs.

Power of SSR-based GDs to Detect EDVs

For fixed thresholds of T=0.25, 0.20, 0.15, and 0.10, substantial differences for the Type I error α and the Type II error $1-\beta_T$ were found between the three material groups. Further analyses revealed that pooling of flint and dent data would lead to a significant increase of flint lines in the fraction of EDVs (data not shown). Moreover, developing a

joint threshold for intra-pool and inter-pool progenies would result in a substantially greater risk of developing an EDV from intra-pool than inter-pool crosses. Consequently, a pool-specific approach is more fair in terms of α and $1-\beta_T$ than fixed GD thresholds. Therefore, thresholds T need to be gene pool specific and different thresholds must be developed for potential EDVs from intra-pool crosses than for progenies from inter-pool crosses.

The thresholds calculated for simulated $GD_{(P1,O)}$ values were generally lower than from observed data. This can be partially explained by the occurrence of non-parental alleles. The most probable reason for this shift, however, is the fact that simulated GD values were based on $\hat{\mu}_{GD_{(P1,P2)}}$ and $\sigma^2_{GD_{(P1,P2)}}$ of all pairwise distances between unrelated (*f*=0) lines within a material group. But breeders often prefer using genetically diverse inbred lines within a gene pool as parents for their recycling breeding. This iplies that the parental lines used in breeding programs may not be a random sample of all unrelated lines of a germplasm pool. When the generation of simulated GD values was repeated with the mean and variance of only the GDs of parental lines actually used, this resulted in a shift of the thresholds towards the corresponding experimental values.

Precision of GDs and Number of Markers Required

Apart from their Type I and Type II errors, the robustness of $GD_{(P1,O)}$ against addition, substitution, or removal of markers is an important factor to be considered for the development of appropriate thresholds. Standard errors (SEs) attached with GD values were of considerable size across all scenarios and material groups, but decreased with decreasing GD thresholds. Assuming a 95% confidence interval (CI) for GD thresholds, this would range from –2 SEs to +2 SEs and, *e.g.*, from 0.13 to 0.29 for $T_{0.05}$ (F_2 *vs.* BC_1 for flint lines) and from 0.04 to 0.14 (BC_1 *vs.* BC_2 for flint lines). Thus, a number of 100 SSRs seems to be at the lower limit for identification of EDVs, as high SEs for GDs increase the probability of Type I or Type II errors. Hence, we recommend a two-stage procedure for identification of EDVs with SSRs in which a set of 100 SSRs uniformly distributed across the genome is analyzed initially, and if there are doubts about the relationship of an IV and a potential EDV, a second set of 100 or more SSR markers is analyzed subsequently.

Given a maximum SE of GDs, one can calculate the necessary number of markers to reach this SEs depending on the mean number of alleles per marker. Minimizing the mean SE for GDs to acceptable values of 0.02 or 0.01 in our study would require a substantial increase in the necessary number of SSRs. For example, a minimum of 260 SSRs would be required to reduce the average SE to 0.01 at a GD level of 0.20. As an alternative, the SE of GDs can be reduced by the choice of highly polymorphic SSR markers. The effective number of alleles (n_e) in our study was 4.2. If it could be doubled to n_e=8.4 by an appropriate choice of highly polymorphic SSR markers only ~120 SSRs would be required to reduce the average SE to 0.01 at a GD level of 0.20. As this high degree of polymorphism is rather unrealistic in connection with an equal distribution of markers over the maize genome, different types of markers seem more promising. This includes the use of a standard DNA chips for EDV identification with the extended large-scale use of expression patterns, the high-throughput application of newly developed marker systems like single nucleotide polymorphisms (SNPs), or the targeted use of allelic information such as the generation of haplotypes.

In addition, intra-varietal variation caused by heterogeneity within lines and lab errors must be considered for the development of thresholds. If ignored, this leads to an overestimation of GDs and, therefore, a bias to the benefit of the breeder of the potential EDV (Heckenberger et al., 2002). Hence, thresholds should be adapted accordingly to reduce this bias. In addition, thresholds must be specific according to a particular set of markers, as the specific choice of markers should be neutral with regard to the conclusion, whether a variety is deemed as an EDV or not.

Appropriate Distance Measures

It is desirable that the GDs between the progeny and either parent add up to the GD between the parental lines (Melchinger, 1993). From all commonly used genetic distance measures, this criterion holds generally only true for the Rogers' (1972) and the Nei and Li (1979) distance. In addition, a linear relationship to *f* is requested, which is fulfilled by both GD measures. Coefficients like Dice (1945), Jaccard (1908), or simple matching (Sneath and Sokal, 1973) are based on single bands, irrespective of the marker to which they belong. Therefore, heterozygous loci are overweighted. In contrast, Rogers' distance

is based on the frequencies of alleles of each marker. Therefore, multiple alleles for a particular marker are weighted in comparison with homozygous alleles of another marker. In addition, frequency-based distance measures could be applied for population varieties and it would be possible to include codominant data from dominant marker systems (Jansen et al., 2001; Piepho and Koch, 2000). Therefore, we recommend the Rogers' distance for identification of EDVs with SSRs. Moreover, we conclude that EDV thresholds must be specific for the distance measure used.

Conclusions

Our results showed that GDs based on SSRs are suitable tools to distinguish between progenies derived from F_2, BC_1 or BC_2 source populations, however, associated with a certain error rate. Due to the observed overlaps in the frequency distributions of $GD_{(P1,O)}$ for F_2-, BC_1-, and BC_2-derived progenies, the choice of an appropriate threshold T is a crucial issue to minimize the Type I (α) and the Type II ($1-\beta_T$) errors. Whereas the GD threshold suggested by ASSINSEL (0.20) results in fairly acceptable α and $1-\beta_T$ values for flint lines, but fairly low $1-\beta_T$ values for dent and introgression lines, we recommend crop- and genepool-specific thresholds on the basis of a fixed α level or $\alpha=1-\beta_T$. Furthermore, the threshold should depend on the marker set and distance measure chosen. Implementation of the EDV concept in practical plant breeding requires a standard set of a large number of highly polymorphic markers for reliable determination of GDs. In addition, we strongly recommend replications of lab assays to minimize lab errors. The frequency distributions of GDs used in this study were based on unrelated parental lines. Obviously, use of related parents for the development of new varieties by recycling breeding will increase the probability of breeding an EDV from an accepted breeding procedure.

Acknowledgements

We are indebted to the 'Gesellschaft zur Förderung der privaten Pflanzenzüchtung' (GFP), Germany for a grant to support Dr. Bohn and the SSR analyses of the present study. Financial support for M. Heckenberger was provided by a grant from the European Union, Grant No. QLK-CT-1999-01499 (MMEDV).

References

ASSINSEL. 1999. Essential derivation and dependence. Practical Information.

ASSINSEL. 2000. DUS testing: phenotype *vs.* genotype. Position Paper adopted at the Rome Congress in May 2000.

Bernardo, R., A. Murigneux, J. P. Maisonneuve, C. Johnsson, and Z. Karaman. 1997. RFLP-based estimates of parental contribution to F_2- and BC_1-derived maize inbreds. Theor. Appl. Genet. 94:652-656.

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

Bohn, M., M. Frisch, and A. E. Melchinger. 2004. A statistical framework for identification of essentially derived varieties obtained from biparental crosses of homozygous lines. Statistical theory. In Preparation.

Dice, L. R. 1945. Measures of the amount of ecologic association between species. Ecology. 26:297-302.

Efron, B. 1979. Bootstrap methods: another look at the jackknife. Ann. Stat. 7:1-26.

Frisch, M., M. Bohn, and A. E. Melchinger. 2000. Plabsim: Software for simulation of marker-assisted backcrossing. J. Hered. 91:86-87.

Grandillo, S., H. M. Ku, and S. D. Tanksley. 1999. Identifying the loci responsible for natural variation in fruit size and shape in tomato. Theor. Appl. Genet. 99:978-987.

Heckenberger, M., M. Bohn, J. S. Ziegle, L. K. Joe, J. D. Hauser, M. Hutton, and A.E. Melchinger. 2002. Variation of DNA fingerprints among accessions within maize inbred lines and implications for identification of essentially derived varieties. I. Genetic and technical sources of variation in SSR data. Mol. Breed. 10:181-191.

Ihaka, R. and R. Gentleman. 1996. R: A language for data analysis and graphics. J. Comp. Graph. Stat. 5:299-314.

International Seed Federation. Chicago, May 2002. ISF view on Intellectual Property.

Jaccard, P. 1908. Nouvelles recherches sur la distribution florale. Bull. Soc. Vaud. Sci. Nat. 44:223-270.

Jansen, R.C., H. Geerlings, A.J. Van Oeveren, and R.C. Van Schaik. 2001. A comment on codominant scoring of AFLP markers. Genetics. 158:925.

Johnson, N. L., S. Kotz, and N. Balakrishnan. 1995. Continuous Univariate Distributions Vol.2. p. 210-275. J. Wiley and Sons, Inc., New York.

Lehmann, E.L. 1986. Testing Statistical Hypotheses. J. Wiley and Sons, New York.

Levene, H. 1960. Robust tests for equality of variances. p. 278-292. *In* Contributions to probability and statistics: Essays in honour of Harold Hotelling. Stanford University Press, Stanford, CA.

Lübberstedt, T., A. E. Melchinger, C. Dussle, M. Vuylsteke, and 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 Sci. 40:783-791.

Malécot, G. 1948. Les Mathematiques de l'Heredité. Masson & Cies, Paris.

Melchinger, A. E., A. Graner, M. Singh, and M. M. Messmer. 1994. Relationships among European barley germplasm. 1. Genetic diversity among winter and spring cultivars revealed by RFLPs. Crop Sci. 34:1191-1199.

Melchinger, A. E. 1993. Use of RFLP markers for analysis of genetic relationships among breeding materials and prediction of hybrid performance. *In* Int.Crop Science I. CSSA. Madison, WI. p.621-628.

Messmer, M. M., A. E. Melchinger, J. Boppenmaier, E. Brunklaus Jung, and R.G. Herrmann. 1992. Relationships among early European maize inbreds. I. Genetic diversity among flint and dent lines revealed by RFLPs. Crop Sci. 32:1301-1309.

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

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

Pejic, I., P. Ajmone-Marsan, M. Morgante, V. Kozumplick, 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.

Piepho, H.-P. and G. Koch, 2000. Codominant analysis of banding data from a dominant marker system by normal mixtures. Genetics. 155:1459.

Rogers, J. S. 1972. Measures of genetic similarity and genetic distance. Studies in Genetics VII. Univ. Texas Publ. 7213:145-153.

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.

Smith J.S.C. and O.S. Smith. 1989. The description and assessment of distances between inbred lines of maize: II. The utility of morphological, biochemical, and genetic descriptors and a scheme for testing of distinctiveness between inbred lines. Maydica. 34:151-161.

Sneath, P. and R. Sokal. 1973. Numerical taxonomia. WH Freeman & Co, San Francisco.

Snedecor, G. and W. Cochran. 1980. Statistical Methods. Iowa State University Press, Ames, IA.

Tivang, J. G., J. Nienhuis, and O. S.Smith. 1994. Estimation of sampling variance of molecular marker data using the bootstrap procedure. Theor. Appl. Genet. 89:259-264.

Troyer, A. F. and T. R. Rocheford. 2002. Germplasm ownership: related corn inbreds. Crop Sci. 42:3-11.

UPOV. 1978. International convention for the protection of new varieties of plants.

UPOV. 1991. International convention for the protection of new varieties of plants.

Wang, J. K. and R. Bernardo. 2000. Variance of marker estimates of parental contribution to F_2- and BC_1-derived inbreds. Crop Sci. 40:659-665.

Tables & Figures

Table 1. Number and type of parent-offspring triplets

	Dent	Flint	Introgression	Total
F_2	38	68	29	136
$BC_{1/2}$	12	0	16	27
Total	50	68	45	163

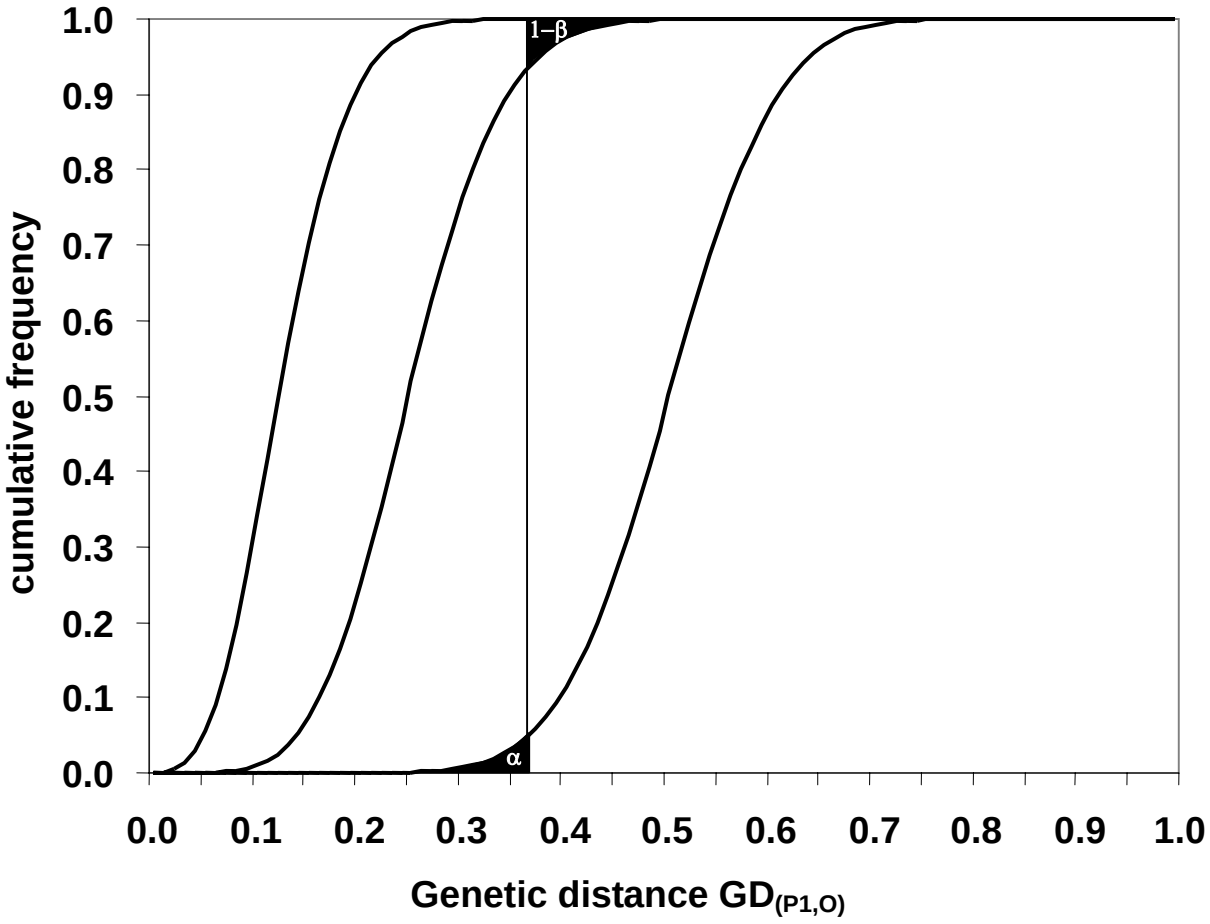


Figure 1. Cumulative frequency distributions of $GD_{(P1,O)}$ for F_2-, BC_1-, and BC_2-derived progenies based on simulated data assuming $GD_{(P1,P2)}$=0.00 Type I (α) and Type II (1-β) errors refer to T=0.37 for discriminating F_2 *vs.* BC_1-derived progeny lines.

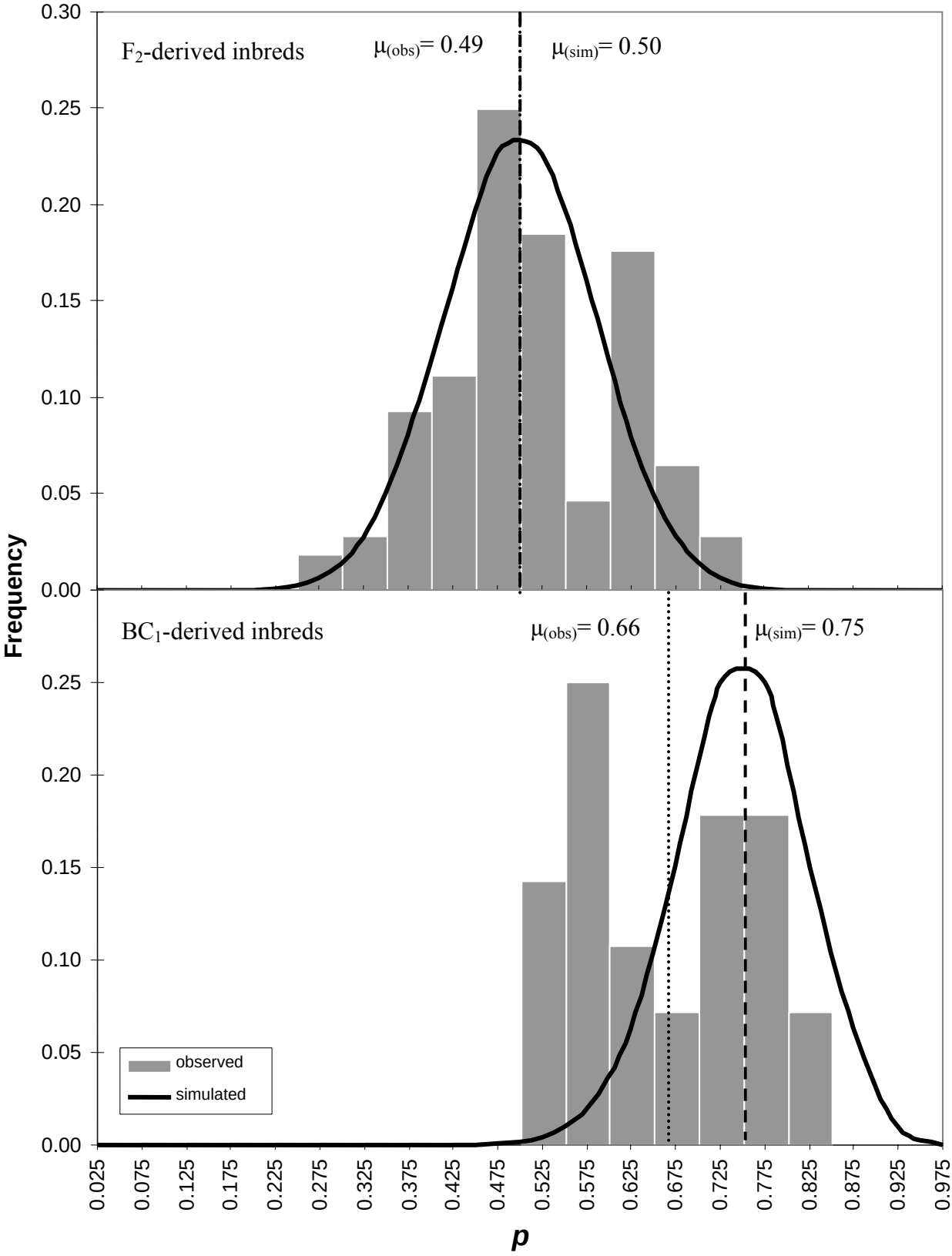


Figure 2. Histograms of observed (columns) and simulated (curve) parental contributions (*p*) for F_2- and BC_1-derived progeny lines. Observed ($\mu_{(obs)}$) and simulated ($\mu_{(sim)}$) means are indicated by dotted and dashed lines, respectively.

Table 2. Comparison of observed and simulated or calculated parameters for the statistical concept used in this study.

Parameter	F_2		BC_1	
	observed	simulated	observed	simulated
μ_p	0.4893	0.5000	0.6567	0.7500
σ_p^2	0.0107	0.0102	0.0088	0.0076
$\mu_{GD_{(P1,P2)}}$	0.6314		0.7277	
$\sigma^2_{GD_{(P1,P2)}}$	0.0024		0.0034	
$\hat{\mu}_{GD_{(P1,O)}}$	0.3095	0.3157†	0.2465	0.1819†
$\sigma^2_{GD_{(P1,O)}}$	0.0051	0.0063‡	0.0034	0.0043‡

† Calculated with Eq. [1] on the basis of observed $\mu_{GD_{(P1,P2)}}$ and $\sigma^2_{GD_{(P1,P2)}}$ and simulated μ_p and σ_p^2

‡ Calculated with Eq. [4] on the basis of observed $\mu_{GD_{(P1,P2)}}$ and $\sigma^2_{GD_{(P1,P2)}}$ and simulated μ_p and σ_p^2

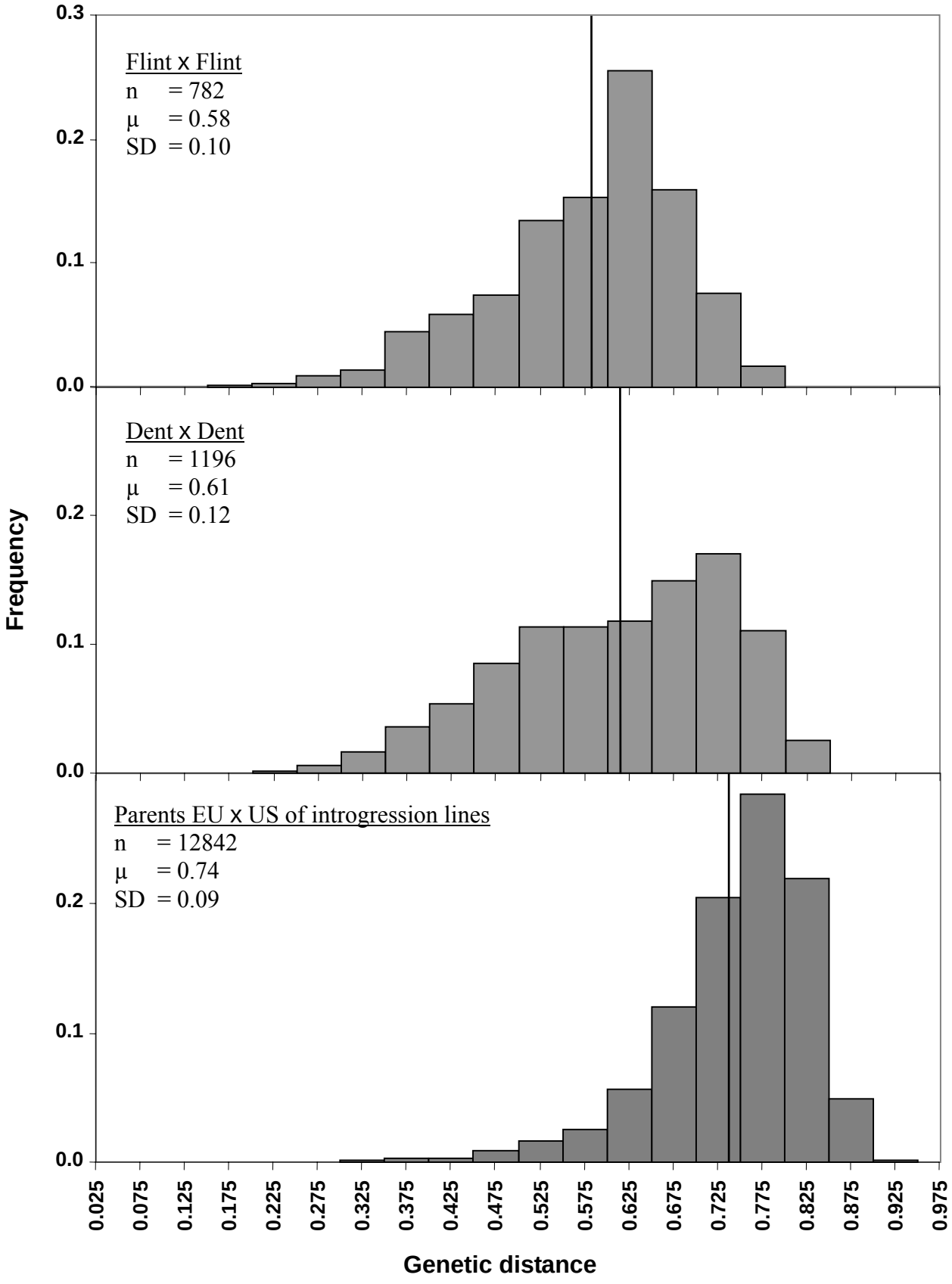


Figure 3. Histograms of Rogers' genetic distances calculated from SSR data between unrelated (*f*=0) lines from European Flint and Dent and parents of introgression lines. Means are indicated by solid lines.

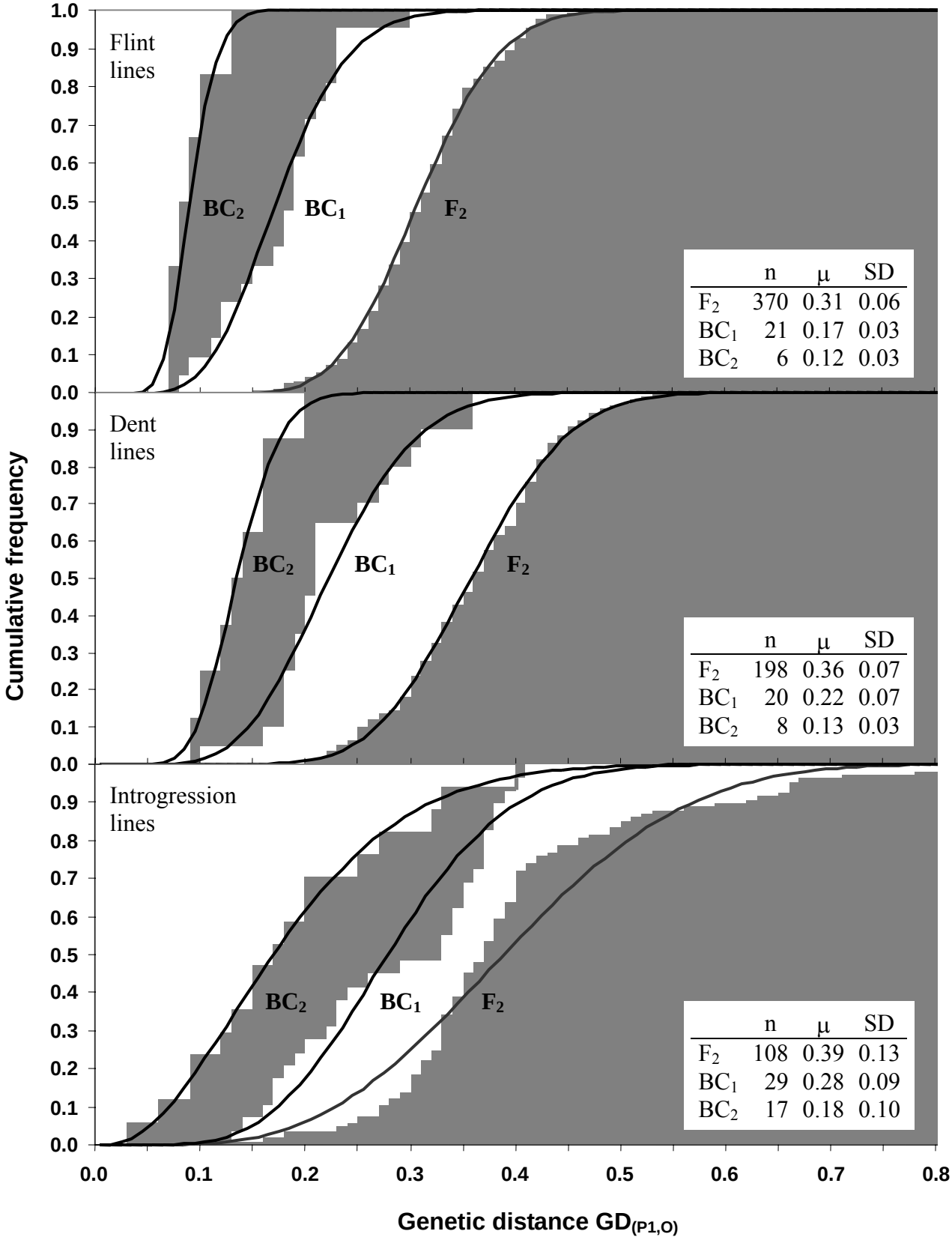


Figure 4. Cumulative histograms (columns) and approximated beta distributions (curves) for genetic distances based on 100 SSRs for F_2-, BC_1-, and BC_2-derived progenies. Variables n, μ, and SD refer to the number of values, the arithmethic mean, and the standard deviation of GD values for the particular distribution, respectively.

Table 3. Evaluation of the discriminatory power of different EDV thresholds (T) based on observed and simulated data. $\overline{SE}$ refers to the average standard error of observed GDs, which were equal to the particular threshold.

Scenario	Flint lines		Dent lines		Introgression lines	
	observed	simulated	observed	simulated	observed	simulated
----------(F$_2$ *vs.* BC$_1$)----------						
***α*=0.05**						
$T_{0.05}$	0.21	0.17	0.24	0.18	0.19	0.24
$(1-\beta_T)$	0.77	0.72	0.63	0.71	0.15	0.85
$\overline{SE}$	0.041	0.036	0.041	0.039	0.039	0.041
α=β						
$T_{\alpha=\beta}$	0.24	0.20	0.28	0.22	0.32	0.26
α=β	0.12	0.12	0.17	0.14	0.32	0.10
$\overline{SE}$	0.041	0.039	0.045	0.042	0.047	0.043
T=0.25 (ASTA)						
α	0.18	0.30	0.07	0.26	0.15	0.07
$(1-\beta_T)$	0.92	0.98	0.68	0.95	0.40	0.90
$\overline{SE}$	0.043	0.043	0.043	0.043	0.043	0.043
***T=0.20* (ASSINSEL)**						
α	0.03	0.11	0.01	0.09	0.07	0.02
$(1-\beta_T)$	0.72	0.87	0.39	0.81	0.19	0.70
$\overline{SE}$	0.039	0.039	0.039	0.039	0.039	0.039
T=0.15 (SEPROMA-orange)						
α	0.00	0.02	0.00	0.02	0.01	0.02
$(1-\beta_T)$	0.36	0.59	0.13	0.53	0.07	0.06
$\overline{SE}$	0.034	0.034	0.034	0.034	0.034	0.034
----------(BC$_1$ *vs.* BC$_2$)----------						
***α*=0.05**						
$T_{0.05}$	0.09	0.07	0.12	0.07	0.15	0.08
$(1-\beta_T)$	0.60	0.55	0.40	0.50	0.29	0.54
$\overline{SE}$	0.027	0.024	0.030	0.027	0.034	0.027
α=β						
$T_{\alpha=\beta}$	0.11	0.10	0.16	0.10	0.24	0.12
α=β	0.11	0.20	0.18	0.18	0.35	0.18
$\overline{SE}$	0.031	0.028	0.036	0.028	0.041	0.030
T=0.15 (SEPROMA-orange)						
α	0.37	0.59	0.13	0.53	0.07	0.06
$(1-\beta_T)$	0.99	0.98	0.72	0.97	0.47	0.44
$\overline{SE}$	0.034	0.034	0.034	0.034	0.034	0.034
T=0.10 (Troyer et al./SEPROMA-red)						
α	0.07	0.00	0.02	0.00	0.01	0.10
$(1-\beta_T)$	0.75	0.20	0.16	0.18	0.10	0.72
$\overline{SE}$	0.028	0.028	0.028	0.028	0.028	0.028

5. Identification of essentially derived varieties (EDVs) derived from biparental crosses of homozygous lines. II. Morphological distances and heterosis in comparison with SSR and AFLP data in maize

M. Heckenberger, M. Bohn, D. Klein, and A. E. Melchinger

Abstract

Morphological traits and heterosis have been proposed apart from genetic distances (GDs) based on molecular markers as possible tools to assess the genetic conformity between putative essentially derived varieties (EDVs) and their initial varieties (IVs). However, for maize and other crops no consensus has been reached regarding methods and thresholds for identification of EDVs, because reliable benchmark data are lacking. The objectives of this study were to (1) evaluate the power of morphological traits and heterosis to discriminate between homozygous progenies derived from F_2, BC_1, and BC_2 populations, (2) compare the findings to published data based on SSRs and AFLPs, and (3) draw conclusions about the usefulness of the various distance measures for identification of EDVs. Morphological distances (MDs) based on 25 traits and mid-parent heterosis for 12 traits were observed for a total of 58 European maize inbred lines comprising 38 triplets. A triplet consisted of one homozygous line derived from a F_2, BC_1 or BC_2 population and both parental inbreds. In addition, all inbreds were genotyped with 100 uniformly distributed SSRs and 20 AFLP primer combinations in companion studies for calculation of genetic distances (GDs). Correlations between the coancestry coefficient, GDs, MDs, and midparent heterosis were significant and high for most traits. However, thresholds for EDVs to discriminate between F_2- and BC_1-derived or BC_1- and BC_2-derived progenies using only morphological distances or heterosis yielded considerably higher values for Type I (α) and Type II ($1-\beta$) errors than observed with GDs based on SSRs and AFLPs. Consequently, we suggest a multi-stage procedure with the initial use of morphological data and a consecutive fingerprinting with molecular markers for identification of EDVs.

Introduction

Plant variety protection (PVP) systems and their laws and regulations should balance commercial interests and warrant sustainable development of new cultivars. On the one hand, registered plant varieties need to be protected against plagiarism and misuse. On the other hand, protected germplasm should be accessible to secure future breeding progress. Therefore, the concept of "breeder's exemption" was introduced into the UPOV convention to solve the obvious conflict between the different stakeholders within the PVP system (UPOV, 1978). Accordingly, plant breeders have access to protected germplasm for the development of new varieties.

New methods such as doubled haploids, marker-assisted backcrossing, and genetic engineering have provided the technical basis to undermine the breeder's exemption in the original sense of the UPOV convention. These tools allow to add a small number of genes to a protected variety and apply for PVP for this "new" variety. In addition, it is possible to select on purpose for lines that are similar to their parents. Therefore, the efforts invested in breeding the original variety can be exploited by the breeder of the plagiarized variety without indemnification. For this reason, the concept of essentially derived varieties (EDVs) was implemented into the revised UPOV convention (UPOV, 1991) and several national plant variety protection acts.

Accordingly, a variety is deemed to be essentially derived from an initial variety (IV), if it is clearly distinguishable but genetically conform to the IV. If the extent of conformity exceeds a certain threshold, the concept of essentially derived varieties (EDVs) indicates that the breeder of the EDV has to arrive at an agreement with the breeder of the IV. However, no consensus has currently been reached on the methods for determining the genetic conformity to distinguish between EDVs and independently derived varieties (IDVs). In addition, accepted or non-accepted breeding procedures have not yet been defined.

Molecular markers, especially simple sequence repeats (SSRs) and amplified fragment length polymorphisms (AFLPs) have been recommended as appropriate tools for determining EDVs in various crops including maize (Dillmann et al., 1997; Bernardo and

Kahler, 2001; Roldan Ruiz et al., 2000a). By contrast, the use of morphological traits or heterosis is still under debate (ASSINSEL, 1999). Hitherto, accurate morphological and agronomic descriptions of cultivars and varieties are the basis of tests for distinctness, uniformity and stability (DUS) within world-wide PVP systems and assure farmers and breeders that they are using clearly identifiable varieties to high standards of purity and quality (Smith and Smith, 1989a). In addition, numerous studies showed significant correlations between midparent heterosis and the coefficient of parentage (*f*) (Melchinger, 1999; Smith et al., 1991). For these reasons, proponents of the use of morphological traits or heterosis for identifications of EDVs argue that phenotypic information provides the basis for PVP and should also be used for identification of EDVs. First studies on the ability of morphological traits to estimate the genetic conformity between related ryegrass (*Lolium perenne* L.) varieties were performed by Gilliland et al. (2000) and Roldan Ruiz et al. (2000b) but revealed only a limited power to distinguish between IDVs and EDVs.

In maize, a triangular instead of a linear relationship was observed between morphological distances and genetic distances or the coancestry coefficient (*f*) (Dillmann and Guérin, 1998). In addition, genetic relationships among maize inbred lines on the basis of morphology were essentially random compared to any relation derived from heterosis or pedigree data (Smith and Smith, 1989b). However, data on the usefulness of heterosis or morphological traits that reflect the degree of relatedness of maize inbred lines in terms of essential derivation is scanty.

The main goal of this study was to investigate the relationship of homozygous progeny lines in maize derived from F_2, BC_1, or BC_2 populations to their parental inbreds based on heterosis and morphological distances (MDs) in comparison with SSR- and AFLP-based genetic distances (GDs). In detail, our objectives were to (1) evaluate the power of heterosis and MDs to discriminate between progenies derived from F_2, BC_1, and BC_2 populations, (2) compare the findings to published data based on SSRs and AFLPs, and (3) draw conclusions about the usefulness of the various distance measures for identification of EDVs.

Materials and Methods

Plant Materials

A total of 58 elite maize inbred lines were analyzed comprising 24 European flint and 34 European dent lines. These lines originated from the maize breeding programs at the University of Hohenheim (Stuttgart, Germany) and two commercial breeding companies in Germany. The 58 lines comprised 38 triplets. A triplet consisted of one progeny line O and both parental lines P1 and P2. The materials consisted of 26 intra-pool triplets of European dent and 12 intra-pool triplets of European flint lines. Progenies were either derived from F_2, BC_1 or BC_2 populations.

For each combination of lines within a triplet (P1xP2, P1xO, and P2xO), seeds from the corresponding F_1 hybrid were generated. In addition, if more than one progeny line (O_1, O_2, …., O_j) was derived from a cross of the same two parental lines, each possible F_1 hybrid (O_1xO_2, …, $O_{j-1}xO_j$) was generated. In total 114 intra-pool F_1 hybrids were tested in this study. Detailed information on all 38 triplets, the 58 maize inbreds and the hybrids included in this study is available as supplemental data in Tables C and D in the appendix of this thesis.

Molecular Analyses

All lines were genotyped with a set of 100 SSR markers uniformly covering the entire maize genome as described in detail by Heckenberger et al. (2002). The 100 SSRs were selected on the basis of reliable single-locus amplification, absence of null alleles, high degree of polymorphism, and high reproducibility of the bands. SSR analyses were performed on a commercial basis by Celera (1756 Picasso Avenue, Davis CA 95616, USA). In addition, all lines were genotyped for AFLPs by Keygene N.V. (P.O. Box 216, 6700 AE Wageningen, The Netherlands). A total of 20 AFLP primer combinations (PCs) was used as described in detail by Heckenberger et al. (2003). AFLP markers were referred to a proprietary integrated map of maize as described by Peleman et al. (2000).

Experimental Design

Field experiments were conducted in 2000 and 2001 at three locations in South Germany with two replications per location. All sites (Bad Krozingen, Eckartsweier, and Scherzheim) are located in the Upper Rhine Valley, a major area of grain-maize production in Germany. All inbred lines and hybrids of a triplet were grown together in one block. Within each triplet block, F_1 hybrids were grown side-by-side with their parental lines to guarantee heterosis estimates with high accuracy. All trials received standard cultural practices of fertilization as well as control of insects and weeds.

The experimental unit was a three-row plot with a row spacing of 0.75 m and a plot length of 4.0 m. Trials were overplanted and later thinned manually to 26 plants per row with a final plant density of 8.7 plants/m^2. Each row was harvested separately. To reduce neighbor effects between adjacent plots with different vigor (inbreds *vs.* hybrids), only data of the middle row of each plot were used for further analyses. The experiment was performed using a randomized block design. Parameter values were observed for 23 morphological traits according to the UPOV guidelines (UPOV, 1978) and 6 additional agronomic traits (Tab. 1) by measuring a minimum of 5 individual plants of a particular plot or by visual observation of the whole plot.

Statistical Analyses

Grain yield for each single row was adjusted to 84.5% dry matter content (DMC). Heterosis was determined as midparent-heterosis $MPH = (F_1 - MP)/MP$, where F_1 is the F_1 hybrid performance and $MP = (P_1 + P_2)/2$ the mid-parent value in which P_1 and P_2 are the performance of the inbred parents, respectively. Analyses of variance (ANOVA) were performed for morphological traits and midparent heterosis using a statistical model considering genotypes as fixed effects and environments as random effects. Heritabilities (h^2) were estimated on an entry-mean basis for all traits. Likewise, heritabilities (h^2_{MPH}) was calculated on a triplet-mean basis for heterotic traits (Hallauer and Miranda, 1981). For calculation of morphological distances (MDs), observations for each trait were standardized by dividing with the phenotypic standard deviation of the particular trait. Euclidean (MD_{EUC}) and Mahalanobis (1936) (MD_{MAH}) distances were calculated based on standardized observations for each pairwise comparison of inbred lines. Malécot's (1948)

coancestry coefficient (*f*) was calculated between all pairwise line combinations. Genetic distances (GDs) between lines based on SSR (GD_{SSR}) or AFLP (GD_{AFLP}) data were estimated using Rogers' distance (1972). The linear relationship between 1-*f*, GDs, MDs, and heterosis estimates was evaluated with a lack-of-fit test (Snedecor and Cochran, 1980). Empirical and approximated frequency distributions of MD values were compared with a Kolmogorov-Smirnov test (Lehmann, 1986) to check for significant deviations. Simple correlations (*r*) were calculated between 1-*f*, GDs, MDs, and heterosis estimates. Homogeneity of variance components of data from flint and dent germplasm was evaluated with Levene's test (1960). Variance components and correlations were not significantly different between flint and dent lines. Consequently, only results from pooled data were reported.

In order to evaluate potential EDV thresholds, the cumulative frequency distributions for genetic distances were approximated with beta distributions (Johnson et al., 1995) as described in detail by Heckenberger et al. (2004a) Frequency distributions for morphological distances and midparent heterosis for F_2-, BC_1-, or BC_2-derived progeny lines were approximated by normal distributions with parameters chosen such that the mean and variance of the original distribution were conserved. Based on these distributions, we calculated Type I (α) and Type II (1-β) errors for various EDV thresholds and various types of populations as suggested in a companion paper for molecular marker data (Heckenberger et al., 2004a). Here, α corresponds to the probability that a true IDV will be wrongly judged as EDV and 1-β corresponds to the probability that a true EDV will not be recognized as such and judged as IDV. We first investigated the situation that an F_2-derived progeny will be considered as IDV, but a BC_1-derived progeny as EDV. Alternatively, we regarded a BC_1-derived progeny as IDV, but a BC_2-derived progeny as EDV.

Statistical analyses of marker data and *f* values were performed as described by Heckenberger et al. (2004a) using the PLABSIM software package (Frisch et al., 2000). ANOVA for field experiments were calculated with the PLABSTAT software (Utz, 2001). All other statistical calculations were carried out with the R software package (Ihaka and Gentleman, 1996).

Results

Morphological Traits and Heterosis Data

Estimates of genotypic variances ($\hat{\sigma}_g^2$) pooled across flint and dent inbred lines were significant (P<0.01) for all traits (Table 1). In addition, significant (P<0.01) genotype x environment interactions were observed for most traits due to cool and wet weather conditions in 2000 and hot and dry weather conditions in 2001. In most cases $\hat{\sigma}_{ge}^2$ was considerably smaller than $\hat{\sigma}_g^2$.

Significant (P<0.01) estimates of σ_g^2 among triplets for MPH were observed for most traits (Table 2). However, considerable differences were found between traits depending on the relative amount of MPH with highest values for grain yield (GYD), grain yield of hand harvested ears (GYE), number of kernels per ear (NKE), and plant length (PLG). Heritabilities for MPH (h^2_{MPH}) of heterotic traits ranged from 0.66 to 0.97 and were slightly smaller than for line *per se* performance (h^2).

SSR and AFLP Marker Data

A total of 580 SSR alleles and 1053 polymorphic AFLP bands was identified for the set of 58 maize lines. The number of alleles per SSR ranged from 3 to 12 with a mean of 5.9. PIC values for SSRs varied between 0.08 and 0.86 and averaged 0.64. The number of polymorphic bands per AFLP primer combination varied from 40 to 70 with an average of 54. PIC values for individual AFLP bands ranged from 0.03 to 0.50 with a mean of 0.33. A detailed description of the genetic diversity revealed by SSRs and AFLPs in this set of lines is given elsewhere.

Relationships Among Distance Measures, Heterosis, and Coancestry

Correlations (r) between 1-f and genetic distances based on SSRs (GD_{SSR}) and AFLPs (GD_{AFLP}) were highly significant (P<0.01) and exceeded 0.85 in both flint and dent lines with a single exception (Table 3). By comparison, r values between GDs and MDs

were medium ($0.40 \leq r \leq 0.68$). Likewise, *r* values between MD_{EUC} and MD_{MAH} were only of moderate size. Correlations for flint lines were consistently higher than for dent lines. Coancestry was moderately correlated with MD_{EUC}, but poorly correlated with MD_{MAH}; both relationships showed a triangular form (Fig. 1).

In contrast, the relationships of GD_{SSR}, GD_{AFLP} and 1-*f* with MPH were linear for most heterotic traits (Fig. 2). Corresponding *r* values were highly significant ($P<0.01$) and moderate to high depending on the trait (Table 2). In general, these correlations were considerably higher than those of MD_{EUC} or MD_{MAH} with GD_{SSR}, GD_{AFLP}, or 1-*f*.

Threshold Scenarios for Identification of EDVs

Observed frequency distributions of MD_{EUC} and MD_{MAH} for F_2-, BC_1-, and BC_2-derived progenies fitted well the approximated normal distributions in the joint analysis of flint and dent lines. Considerable overlaps between the frequency distributions of MD_{EUC} and MD_{MAH} were observed for F_2- *vs.* BC_1- as well as for BC_1- *vs.* BC_2-derived progenies (Fig. 3).

For thresholds based on MDs, the power β to classify a BC_1-derived progeny line as EDV amounted to 18% for MD_{EUC} and 3% for MD_{MAH}, when choosing $\alpha=0.05$ for F_2-derived lines (Table 4). Assuming $\alpha=0.05$ for BC_1-derived lines, corresponding values of β for BC_2-derived lines were considerably higher for MD_{EUC} and MD_{MAH}. The power β for thresholds determined by $\alpha=1-\beta$ to classify BC_1- or BC_2-derived progenies as EDVs increased considerably compared to the values for $\alpha=0.05$. This increase in β was associated with higher values for α. When potential thresholds were based on MPH, the power β to classify a BC_1-derived progeny line as EDV ranged from 2% to 30% assuming $\alpha=0.05$ between F_2-derived lines (Table 4). Choosing $\alpha=0.05$ for BC_1-derived lines, the values of β increased for BC_2-derived lines. For $\alpha=1-\beta$, the power to classify BC_1- or BC_2-derived progenies as EDVs increased substantially, however, this was again associated with higher values for α. In general, values for α and 1-β were of similar magnitude for MDs and MPH.

A detailed description of threshold scenarios based on SSR- and AFLP-based GDs is given in companion papers (Heckenberger et al., 2004a; 2004b).

Discussion

The maize inbred lines examined in our study represent a cross-section of modern elite flint and dent inbred lines from commercial and public maize breeding programs in Germany. Morphological traits were chosen according to the UPOV guidelines of distinctness, uniformity and stability (DUS). In addition, heterosis and morphological traits were determined in extensive field trials over two years and three locations, which exceeds by far the number of environments employed for DUS testing within PVP systems. Furthermore, SSRs were selected as a suitable marker system due to their known map positions and high degree of polymorphism. AFLPs were chosen due to the greater number of markers per assay unit and their high reproducibility (Heckenberger et al., 2003). Thus, the present study is the first larger investigation after a series of pioneering papers based on isozymes and RFLPs (Smith and Smith, 1989a; 1989b; Smith et al., 1991) to provide critical data on the ability of morphological distances and heterosis for identification of EDVs in maize in direct comparison with SSR and AFLP data. For this reason, our results provide a well-founded comparison of different distance measures for identification of EDVs in maize and may serve as an example for other crops.

Data Quality and Relatedness of Different Measures for Genetic Conformity

Despite the contrasting climatic conditions during the vegetation seasons in 2000 and 2001, high heritabilities were observed for morphological traits and midparent heterosis, the former being considerably higher than those reported by Rebourg et al. (2001). In addition, UPGMA cluster analysis based on MD_{EUC} showed a clear grouping of flint and dent lines, which further corroborates the high quality of morphological data (available as supplemental data in Figure A in the appendix of this thesis). The dendrogram based on MD_{MAH} showed as well a grouping of flint and dent lines, but contained several inbreds that were clustered together with lines of the opposite pool. In addition, only moderate correlations between MC_{EUC} and MD_{MAH} were observed. This can be explained by the different statistical properties of MC_{EUC} and MD_{MAH}, because MD_{MAH} adjusts for the correlations of traits.

The graphs between MDs and GDs (Fig. 1) confirmed the triangular relationship of morphological and genetic distances reported in previous studies (Dillmann et al., 1997; Rebourg et al., 2001). This indicates that low GDs correspond necessarily with low MDs, whereas the reverse does not necessarily hold true because high GDs can correspond with both high and low MDs (Van Eeuwijk and Baril, 2001). In addition, the triangular shape has several biological explanations (Nuel et al., 2001) and is also expected, if only molecular markers tightly linked with the genes controlling the phenotypic trait(s) were used (Burstin and Charcosset, 1997).

In the present investigation, flint and dent lines and triplets showed similar estimates of $\hat{\sigma}_g^2$ and correlations among the various criteria. This is in harmony with a previous study of Bar-Hen et al. (1995), who examined 974 maize inbred lines with morphological traits and RFLPs. Correlations of 1-*f*, GD_{SSR}, or GD_{AFLP}, with MPH were higher than reported by Ajmone Marsan et al. (1998) for AFLPs but similar to correlations of MPH with 1-*f* and GDs based on RFLPs in intra-pool crosses (Boppenmaier et al., 1993; Smith et al., 1990). In addition, our study confirms the findings of Smith and Smith (1989b) that correlations of molecular markers or 1-*f* were considerably higher with MPH than with MD_{EUC} or MD_{MAH}.

Distinctness *vs.* Conformity

To confirm an essential derivation in the intention of the UPOV convention (UPOV, 1991), three separate criteria must be fulfilled. An EDV must be (i) distinct from the IV, (ii) 'predominantly derived' from the IV, and (iii) conform to it in the expression of it's 'essential characteristics'. Distinctness can be determined based on morphological traits by established procedures for DUS testing. Establishing a predominant derivation will either require a directly documented evidence, *e.g.*, by breeding books, or could be determined with molecular evidence similar to forensic approaches in the human sector (Gill et al., 1995). However, the question of whether conformity in the expression of essential characteristics should be assessed by phenotypic rather than molecular data is still unsolved. While differences in the expression of one single trait are sufficient to prove distinctness between two varieties, assessment of conformity should be based on a large number of

morphological traits and still could not give a definite answer due to the triangular relationship mentioned above.

Proponents of phenotypic data argue that the term 'conform in the expression of it's essential characteristics that result from the genotype' (UPOV, 1991) implies the use of phenotypic data rather than molecular data. In contrast, opponents state that even highly heritable phenotypic traits can only offer an indirect measure of the relatedness of two cultivars. In contrast, molecular data provide a direct estimate of the true relatedness of two genotypes because it is unbiased from environmental effects and reflects the percentage of the genome in common between the IV and a putative EDV. Based on our results, genetic distances based on molecular markers have clear advantages for identification of EDVs.

Power of Morphological Distances and Heterosis for Identification of EDVs

For MD_{EUC} as well as for MD_{MAH}, extensive overlaps of the frequency distributions of F_2-, BC_1-, and BC_2-derived progenies were found in spite of the significant correlations with 1-*f*. Thus, Type I (α) and Type II (1-β) errors observed for MDs were considerably higher than observed for GDs based on molecular markers (Table 3). Consequently, MDs provide only a rough estimate of the true relatedness of two lines and can only poorly discriminate F_2-, BC_1-, and BC_2-derived progenies. These results confirm data from ryegrass (Gilliland et al., 2000) and maize (Smith and Smith, 1989b) showing that morphological conformity could give an initial indication of the relatedness of two cultivars, particularly for highly conforming pairs of inbreds. However, a small MD between two varieties cannot be taken as a definitive proof that they are in fact closely related because of the triangular relationship between 1-*f* and MDs.

In contrast to the triangular relationship between GDs and MDs, a linear relationship of MPH with GDs or 1-*f* was observed as expected by quantitative genetic theory (Melchinger, 1999). However, in spite of the higher correlation of MPH with 1-*f* or GDs, MPH was not markedly superior to MDs regarding the power to discriminate between F_2-, BC_1- or BC_2-derived progeny. This is attributable to the larger experimental error and GxE interactions of MPH in comparison with line *per se* performance (data not shown) as reflected by the comparison of h^2_{MPH} vs. h^2.

For nearly all scenarios examined, GDs based on SSRs or AFLPs were superior over MDs or MPH for any trait or combination of traits in their power β to discriminate among F_2-, BC_1, and BC_2-derived progenies for given values of α. However, different from MDs and MPH, the use of SSR or AFLP markers would require thresholds specific for a given germplasm pool. This is necessary because flint and dent lines differed significantly in their mean GD between unrelated lines due to the different levels of polymorphism within each germplasm pool (Heckenberger et al., 2004a)

Conclusions

Based on our results, morphological distances and midparent heterosis can provide only an initial indication for putative EDVs. However, a reliable identification of EDVs by MPH or MDs alone is not possible due to the large overlaps in the frequency distributions of MDs and MPH of F_2-, BC_1-, and BC_2-derived progenies. In addition, MDs and MPH have several disadvantages compared to molecular markers. First, assessment of morphological traits and MPH requires extensive field trials over several years and locations to minimize environmental effects. Therefore, these measurements are more expensive and time consuming than molecular marker analyses. Second, heterosis estimates requires production and testing of hybrids. In addition, reciprocal crosses should be evaluated to minimize the risk of maternal effects (Melchinger et al., 1986). Third, the scoring of morphological traits is to some extent subjective. Therefore, a number of check inbreds must be included in the study to warrant a high quality of morphological traits across different years and scoring persons.

In conclusion, we recommend a multi-stage procedure for identification of EDVs with the initial use of morphological data from DUS testing and a consecutive fingerprinting with a first set of at least 100 SSR markers or 20 AFLP PCs for putative EDVs. If doubts still prevail, whether a variety has been derived independently from another variety or not, the corresponding genotypes should be fingerprinted with a second set of SSRs or AFLPs. Use of MPH for identification of EDVs is problematic, because the rationale for using MPH is merely its linear relationship with 1-*f* and the biological mechanisms underlying heterosis are not fully understood.

Acknowledgements

We are indebted to the 'Gesellschaft zur Förderung der privaten Pflanzenzüchtung' (GFP), Germany, for a grant to support Dr. Bohn and the molecular analyses of the present study. Financial support for M. Heckenberger and the field trials was provided by a grant from the European Union, Grant No. QLK-CT-1999-01499 (MMEDV).

References

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

ASSINSEL. 1999. Essential Derivation and Dependence. Practical Information.

Bar-Hen, A., A. Charcosset, M. Bougoin, and Guiard J. 1995. Relationship between genetic markers and morphological traits in a maize inbred lines collection. Euphytica. 84:145-154.

Bernardo, R. and A. L. Kahler. 2001. North American study on essential derivation in maize: inbreds developed without and with selection from F_2 populations. Theor. Appl. Genet. 102:986-992.

Boppenmaier, J., A. E. Melchinger, G. Seitz, H. H. Geiger, and R. G. Herrmann. 1993. Genetic diversity for RFLPs in European maize inbreds. III. Performance of crosses within versus between heterotic groups for grain traits. Plant Breed. 111:217-226.

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

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

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

Frisch, M., M. Bohn, and A. E. Melchinger. 2000. Plabsim: Software for simulation of marker-assisted backcrossing. J. Hered. 91:86-87.

Gill, P., C. P. Kimpton, A. Urquhart, N. Oldroyd, E. S. Millican, S. K. Watson, and T. J. Downes. 1995. Automated short tandem repeat (STR) analysis in forensic casework: a strategy for the future. Electrophoresis 16:1543-1552.

Gilliland, T. J., R. Coll, E. Calsyn, M. de. Loose, M. J. T. v. Eijk, and I. Roldan Ruiz. 2000. Estimating genetic conformity between related ryegrass (*Lolium*) varieties. 1. Morphology and biochemical characterisation. Mol. Breed. 6:569-580.

Hallauer, A. R. and J.B. Miranda. 1981. Quantitative Genetics in Maize Breeding. The Iowa State University Press, Ames, Iowa.

Heckenberger, M., M. Bohn, and A. E. Melchinger. 2004a. A statistical framework for identification of essentially derived varieties derived from biparental crosses of homozygous lines. II. SSR data from maize. Crop Sci. submitted.

Heckenberger, M., M. Bohn, J. Rouppe van der Voort, and A. E. Melchinger. 2004b. A statistical framework for identification of essentially derived varieties derived from biparental crosses of homozygous lines. IV. AFLP data from maize and comparison with SSR data. In preparation.

Heckenberger, M., M. Bohn, J. S. Ziegle, L. K. Joe, J. D. Hauser, M. Hutton, and A. E. Melchinger. 2002. Variation of DNA fingerprints among accessions within maize inbred lines and implications for identification of essentially derived varieties. I. Genetic and technical sources of variation in SSR data. Mol. Breed. 10:181-191.

Heckenberger, M., J. Rouppe van der Voort, A. E. Melchinger, J. Peleman, and M. Bohn. 2003. Variation of DNA fingerprints among accessions within maize inbred lines and implications for identification of essentially derived varieties. II. Genetic and technical sources of variation in AFLP data and comparison to SSR data. Mol. Breed. 12:97-106.

Ihaka R. and R. Gentleman. 1996. R: A Language for data analysis and graphics. J. Comp. Graph. Stat. 5:299-314.

Johnson, N. L., S. Kotz, and N. Balakrishnan. 1995. Continuous Univariate Distributions Vol.2. p. 210-275. J. Wiley and Sons, Inc., New York.

Lehmann, E. L. 1986. Testing Statistical Hypotheses. John Wiley and Sons, New York.

Levene, H. 1960. Robust tests for equality of variances. p. 278-292. *In* Contributions to probability and statistics: Essays in honour of Harold Hotelling. Stanford University Press, Stanford, CA.

Mahalanobis P.C. 1936. On the generalized distance in statistics. Proc. Natl. Acad. Sci. India. 2:49-45.

Malécot, G. 1948. Les mathematiques de l'heredité. p. Masson & Cies, Paris.

Melchinger, A.E. 1999. Genetic diversity and heterosis. Int. Symp. on Genetics and Exploitation of Heterosis in Crop Plants. p. 99-118. *In* J.G. Coors and S. Pandey (ed.) The genetics and exploitation of heterosis in crops. ASA, CSSA, and SSSA, Madison, WI.

Melchinger, A. E., H. H. Geiger, and F. W. Schnell. 1985. Reciprocal differences in single-cross hybrids and their F_2 and backcross progenies in maize. Maydica. 30:395-405.

Nuel, G., S. Robin, and C.P. Baril. 2001. Predicting distanced using a linear model: the case of varietal distinctness. J. Appl. Stat. 28:607-621.

Peleman, J., R. van Wijk, J. van Oeveren and R. van Schaik. 2000. Linkage map integration:an integrated genetic map of *Zea mays* L. Poster P472. Plant & Animal Genome Conference VIII. San Diego.

Rebourg, C., B. Gouesnard, and A. Charcosset. 2001. Large scale molecular analysis of traditional European maize populations. Relationships with morphological variation. Heredity. 86:574-587.

Rogers, J. S. 1972. Measures of genetic similarity and genetic distance. Studies in Genetics VII. Univ. Texas Publ. 7213:145-153.

Roldan Ruiz, I., E. Calsyn, T. J. Gilliland, R. Coll, M. J. T. van Eijk, and De LooseM. 2000a. Estimating genetic conformity between related ryegrass (*Lolium*) varieties. 2. AFLP characterization. Mol. Breed. 6:593-602.

Roldan Ruiz, I., F. A. v. Eeuwijk, T. J. Gilliland, P. Dubreuil, C. Dillmann, J. Lallemand, M. de. Loose, and C. P. Baril. 2000b. A comparative study of molecular and morphological methods of describing relationships between perennial ryegrass (*Lolium perenne* L.) varieties. Theor. Appl. Genet. 103:1138-1150.

Smith J.S.C. and Smith O.S. 1989a. The description and assessment of distance between inbred lines of maize: I. The use of morphological traits as descriptors. Maydica. 34:141-150.

Smith J.S.C. and Smith O.S. 1989b. The description and assessment of distances between inbred lines of maize: II. The utility of morphological, biochemical, and genetic descriptors and a scheme for testing of distinctiveness between inbred lines. Maydica. 34:151-161.

Smith, J. S. C., O. S. Smith, S. L. Bowen, R. A. Tenborg, and 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 RFLPs. Maydica. 36:213-226.

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, F1 grain yield, grain yield, heterosis, and RFLPs. Theor. Appl. Genet. 80:833-840.

Snedecor, G. and W. Cochran. 1980. Statistical methods. Iowa State University Press, Ames, IA.

UPOV. 1978. International convention for the protection of new varieties of plants.

UPOV. 1991. International convention for the protection of new varieties of plants.

Utz, H. F. 2001. Plabstat, ein Computerprogramm zur statistischen Analyse von pflanzenzüchterischen Experimenten, Version 2P vom 14. Juli 2001 (in German). University of Hohenheim, Germany.

Van Eeuwijk, F. A. and C. P. Baril. 2001. Conceptual and statistical issues related to the use of molecular markers for distinctness and essential derivation. Acta Hort. 546:35-53.

Tables & Figures

Table 1. Morphological and agronomic traits, their genotypic (σ_g^2) and genotype x environment interaction (σ_{ge}^2) variances and heritabilities (h^2) observed for 58 flint and dent maize lines in four or six environments in South Germany.

Trait	code	UPOV code	Test sites	σ_g^2	σ_{ge}^2	$h^{2\dagger}$
Ear						
diameter (mm)	EDI	27	6	9.28**	0.47**	96.2
number of kernels	NKE	-/-	6	3108**	558**	88.6
type of grain (1-9 scale)	TGR	30	4	2.04**	0.09**	97.4
anthocyanin coloration (col.) of glumes of cob	AGC	34	4	4.48**	0.05*	98.4
anthocyanin col. of silks (1-9 scale)	ACS	17	4	1.27**	0.22**	87.7
color of dorsal side of grain (1-9 scale)	CDG	32	4	0.37**	0.02	86.6
color of tip of grain (1-9 scale)	CTG	31	4	1.13**	0.03**	96.5
length (mm)	ELG	26	6	188.9**	39.9**	92.6
number of rows of grain	NGR	29	6	1.19**	0.14**	92.8
days to silk emergence	TSE	15	4	33.1**	3.45**	96.4
Kernels						
thousand kernel weight (g)	TKW	-/-	6	1059**	168**	95.4
grain yield ($Mg*ha^{-1}$)	GYD	-/-	6	107.9**	39.2**	79.4
grain yield of 4 hand harvested ears (g)	GYE	-/-	6	3.72**	0.47**	90.4
Leaf						
angle between blade and stem (1-9 scale)	LAN	4	4	1.04**	0.21**	88.4
attitude of blade (1-9 scale)	LAT	3	4	0.72**	0.02	90.1
width of blade (mm)	WBL	24	4	0.44**	0.11**	81.3
Plant						
ear height (cm)	EHT	-/-	4	164.8**	29.0**	95.4
length (cm)	PLG	22	6	484.0**	51.3**	94.7
Tassel						
anthocyanin col. of base of glume (1-9)	ABG	8	4	2.90**	0.15**	94.7
anthocyanin col. of glume excluding base (1-9)	AEB	9	4	1.41**	0.19**	88.9
length of side branches (br.)(cm)	LSB	21	4	3.90**	1.19**	86.6
angle between main axis and lateral br. (1-9)	TAN	12	4	1.67**	0.25**	88.7
anthocyanin col. of anthers (1-9)	AAH	10	4	1.18**	0.35**	85.0
attitude of lateral branches (1-9)	ALB	13	4	1.59**	0.59**	80.6
length of main axis above lowest side br. (cm)	TLL	19	4	8.16**	2.73**	84.7
length of main axis above upper side br. (cm)	TLU	20	4	12.5**	0.44**	83.2
number of primary lateral branches	NLB	14	4	1.23**	0.18**	91.2
days to anthesis	TAH	7	4	21.7**	1.42**	96.7

$^{\dagger}h^2$ = heritability on an entry-mean basis for line *per se* performance pooled across flint and dent lines

Table 2. Estimates of mean, range, genotypic (σ_g^2) and genotype x environment interaction (σ_{ge}^2) variances, heritability (h^2_{MPH}) of mid-parent heterosis (MPH) observed for different morphological and agronomic traits of 114 flint and dent hybrids and their parental lines tested in four or six environments in South Germany as well as correlations (*r*) of MPH with coancestry (1-*f*), genetic distance based on 100 SSRs (GD_{SSR}) or 20 AFLP primer combinations (GD_{AFLP}), Euclidean (MD_{EUC}) or Mahalanobis (MD_{MAH}) morphological distances.

Trait†	MPH			σ_g^2	σ_{ge}^2	h^2_{MPH}‡	*r*				
	Mean	Min.	Max.				1-*f*	GD_{SSR}	GD_{AFLP}	MD_{EUC}	MD_{MAH}
EDI	0.09	0.01	0.18	< 0.001	< 0.001	91.3	0.68**	0.78**	0.76**	0.65**	0.55**
NKE	0.59	0.03	1.34	0.070**	0.012**	96.3	0.76**	0.86**	0.87**	0.69**	0.60**
ELG	0.25	0.02	0.53	0.010**	< 0.001	96.9	0.78**	0.85**	0.89**	0.68**	0.63**
NGR	0.06	-0.02	0.16	0.010**	0.001**	73.2	0.38**	0.56**	0.50**	0.43**	0.39**
TSE	-0.06	-0.17	0.03	0.001**	0.001**	65.9	-0.70**	-0.66**	-0.70**	-0.58**	-0.51**
TKW	0.10	-0.04	0.42	0.004**	0.002**	90.1	0.66**	0.71**	0.76**	0.55**	0.53**
GYD	0.79	0.14	2.14	0.150**	0.054**	92.1	0.73**	0.84**	0.86**	0.72**	0.61**
GYE	0.75	0.06	1.84	0.130**	0.021**	96.9	0.80**	0.90**	0.92**	0.66**	0.63**
EHT	0.27	0.01	0.63	0.010**	0.005**	83.8	0.75**	0.78**	0.80**	0.58**	0.57**
PLG	0.17	0.01	0.36	0.010**	0.001**	95.2	0.75**	0.85**	0.87**	0.63**	0.55**
TAH	-0.05	-0.17	0.04	< 0.001	< 0.001	72.1	-0.59**	-0.76**	-0.74**	-0.49**	-0.42**

*,** Significant at the 0.05 or 0.01 probability level, respectively
† For abbreviations, see Table 1.
‡ h^2_{MPH} = heritability on a triplet-mean basis for mid-parent heterosis pooled across flint and dent lines.

Table 3. Simple correlations between coancestry coefficient (1-*f*), genetic distances based on 100 SSRs (GD_{SSR}) and 20 AFLP primer combinations (GD_{AFLP}) as well as Euclidean (MD_{EUC}) and Mahalanobis (MD_{MAH}) morphological distances based on 25 traits (see Table 1) for 24 flint (below diagonal) and 34 dent inbreds (above diagonal).

	1-*f*	**GD_{SSR}**	**GD_{AFLP}**	**MD_{EUC}**	**MD_{MAH}**
1-*f*		0.75**	0.85**	0.58**	0.31**
GD_{SSR}	0.88**		0.92**	0.57**	0.40**
GD_{AFLP}	0.88**	0.97**		0.68**	0.40**
MD_{EUC}	0.55**	0.65**	0.65**		0.62**
MD_{MAH}	0.44**	0.49**	0.59**	0.76**	

** Significant at the 0.01 probability level.

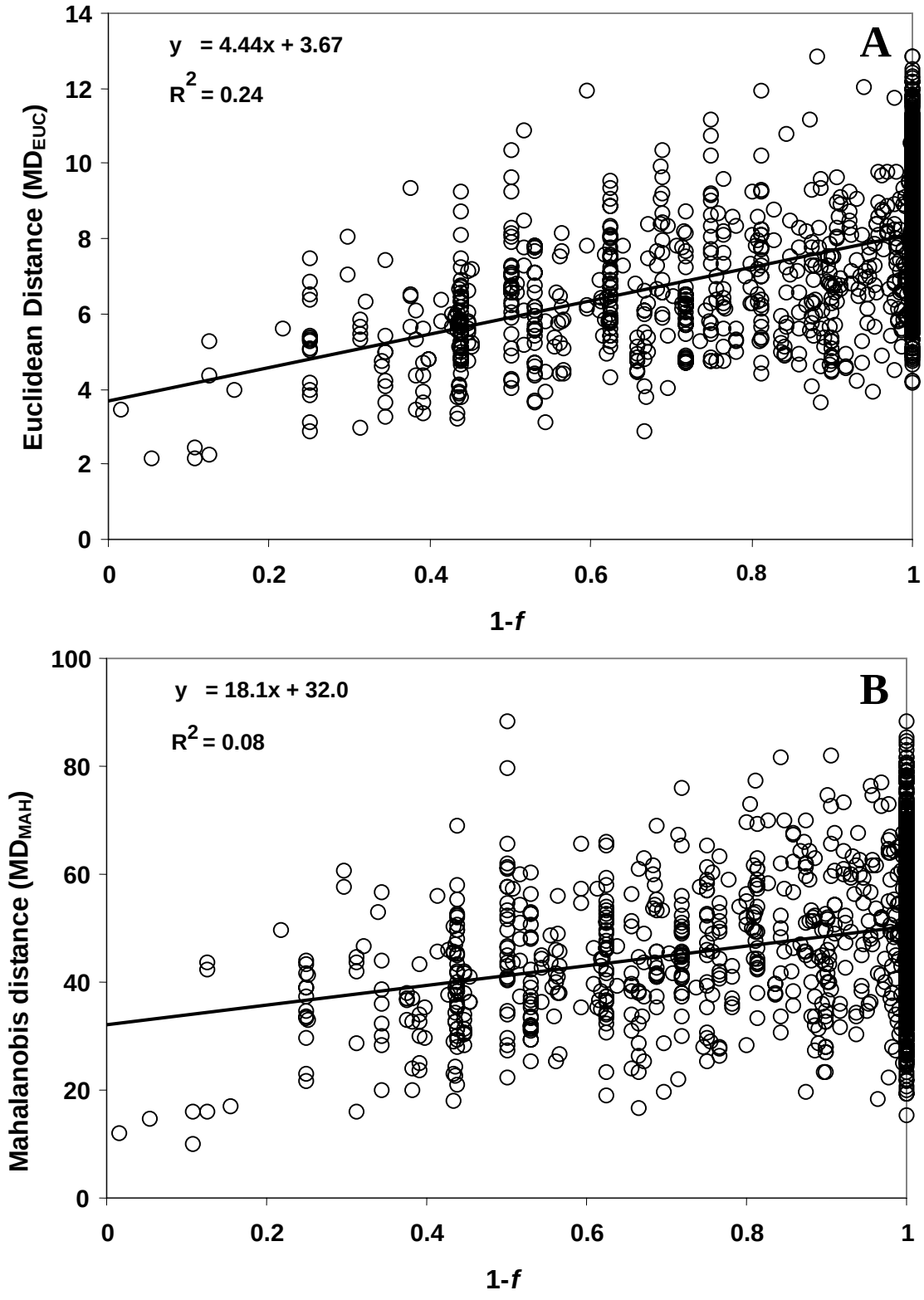


Figure 1. Relationship of the coancestry coefficient (*f*) with (A) Euclidean (MD_{EUC}) and (B) Mahalanobis (MD_{MAH}) distances based on 25 morphological traits observed for 1767 pairwise comparisons of maize inbred lines.

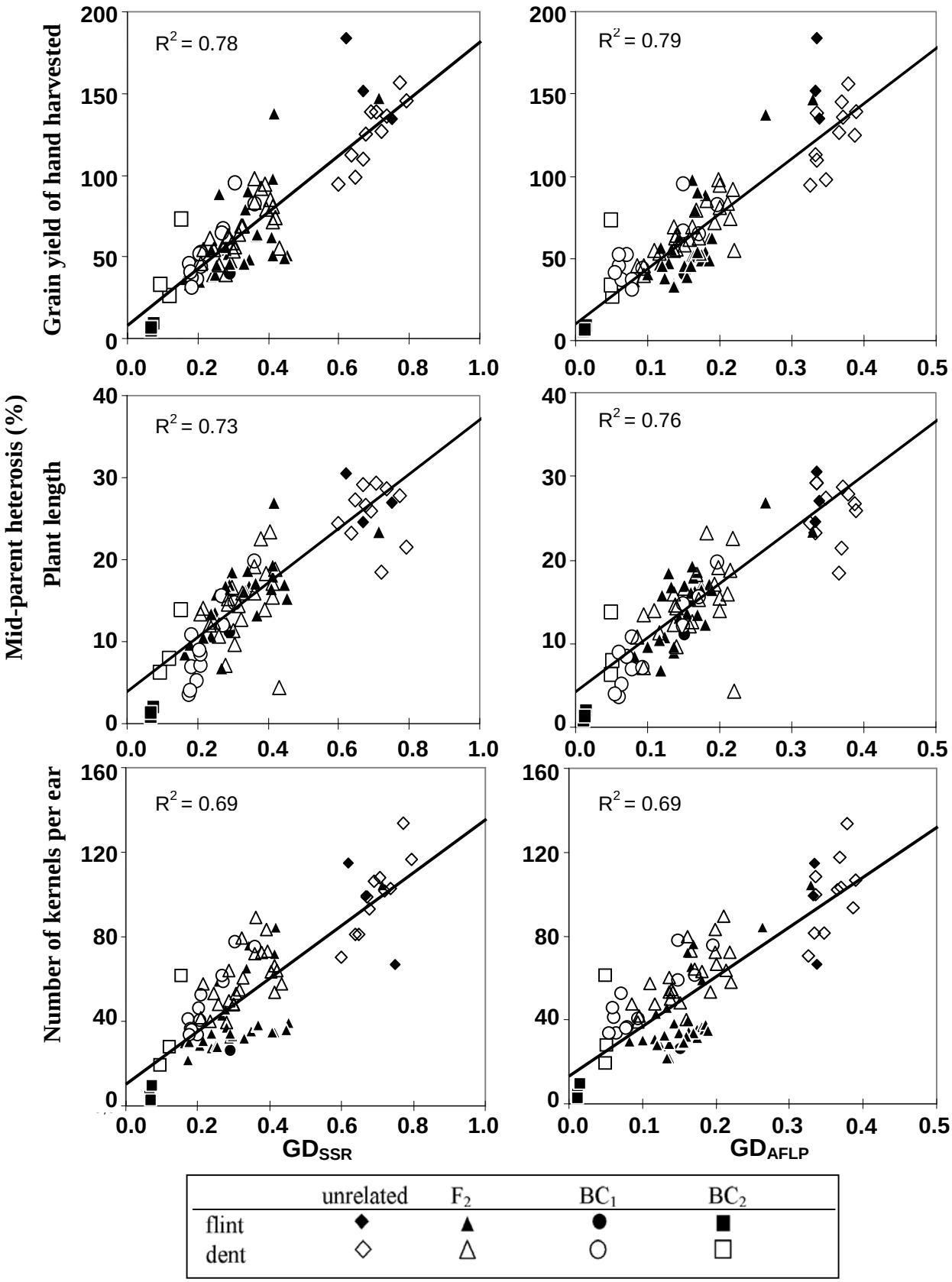


Figure 2. Relationship of genetic distances based on 100 SSRs (GD_{SSR}) or 20 AFLP primer combinations (GD_{AFLP}) with mid-parent heterosis (in %) of 84 intra-pool hybrids with given pedigree relationships of their parental maize lines.

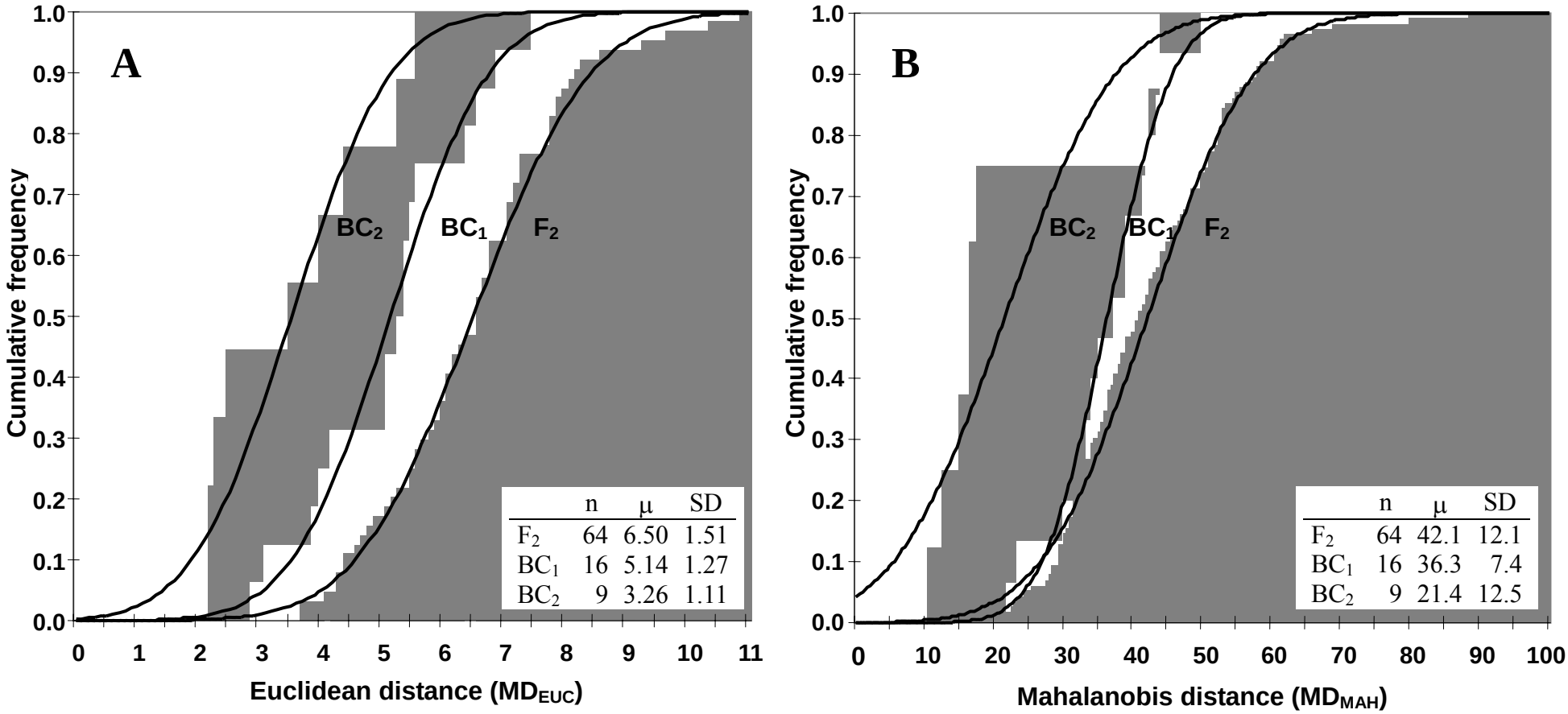


Figure 3. Cumulative histograms (columns) and approximated normal distributions (curves) for (A) Euclidean (MD_{EUC}) or (B) Mahalanobis (MD_{MAH}) morphological distances based on 25 morphological traits for F_2-, BC_1-, and BC_2-derived progeny lines. Variables n, μ, and SD refer to the number of values, the mean, and the standard deviation of MD values for the particular distribution, respectively.

Table 4. Evaluation of different scenarios for thresholds T based on morphological distances, heterosis, and genetic distances based on SSRs and AFLPs.

Parameter	F_2 *vs.* BC_1				BC_1 *vs.* BC_2			
	α=0.05		$\alpha=\beta$		α=0.05		$\alpha=\beta$	
	T	1-β	T	$\alpha = \beta$	T	1-β	T	$\alpha = \beta$
Morphological distances								
Euclidean (MD_{EUC})	4.0	0.18	5.8	0.32	3.1	0.40	4.1	0.21
Mahalanobis (MD_{MAH})	22.5	0.03	38.5	0.39	24.5	0.60	31.0	0.28
Heterosis								
Grain yield (GYE)	0.24	0.05	0.58	0.39	0.24	0.47	0.41	0.25
Plant length (PLG)	0.08	0.29	0.13	0.31	0.03	0.30	0.08	0.29
Number of kernels per ear (NKE)	0.17	0.02	0.48	0.47	0.22	0.52	0.36	0.24
Cumulative[†]	0.14	0.07	0.29	0.38	0.13	0.49	0.21	0.25
Genetic distances								
100 SSRs (GD_{SSR})	0.21	0.68	0.25	0.14	0.08	0.38	0.12	0.18
20 AFLP PCs (GD_{AFLP})	0.12	0.65	0.14	0.21	0.04	0.37	0.05	0.10

[†]Average relative heterosis of five traits (GYE, ELG, NKE, PLG, EHT) showing highest correlation with 1-*f*; for abbreviations, see Table 1.

6. General discussion

The Concept of Essential Derivation within the Tendencies in Modern Plant Breeding

A selection for simple and complex traits to improve domesticated animals or plants in highly developed long-term selection processes has for centuries been performed entirely on their phenotypes. Even though this has proven to be a fabulously successful approach, the forthcoming age of biotechnology and genomics offers the prospect of shifting selection gradually from phenotypes to genotypes (Walsh, 2001). In addition, the available genome sequence of *Arabidopsis thaliana* (The Arabidopsis Genome Initiative, 2000) as well as the growing number of identified genes in major crops, such as *brachytic* (Multani et al., 2003) or *dwarf8* (Thornsberry et al., 2001) in maize (*Zea mays* L.), have provided plant breeders with new perspectives as the exploitation of sequence homologies with other crops or inter-specific introgression of favorable genes are (Walsh, 2000).

In combination with already available tools like marker-assisted selection, these new approaches steadily reduce the time intervals necessary for breeding new varieties. This increases the pressure on plant breeders to release new varieties to keep up with the breeding progress of competitors. Therefore, marker and sequence information, cloned genes, germplasm resources as well as protected germplasm must be available to all breeders, ensuring the most efficient breeding progress to the breeders of all crops. The concept of essential derivation, as implemented in the 1991 Act of the UPOV convention (UPOV, 1991), is thus a first step towards a framework for regulations of the exchange of germplasm among breeders and could be followed by regulations for exchange of marker information or DNA-sequences.

Identification of EDVs and Rating of Threshold Scenarios

Since its implementation in 1991, the EDV concept has gained explicit proponents as well as severe criticism. Troyer and Rocheford (2002) pled for low EDV thresholds ($T \leq 0.10$) with only low royalty fees to be paid by the breeders of EDVs over a short period of time because higher thresholds would cause more EDVs and fewer IDVs. As a result, more EDVs would give rise to more lawsuits and more royalty payments and thus more

money for lawyers and accountants. Investments in the germplasm development would consequently be limited and the rate of yield improvement would be expected to decrease (Duvick, 1984). Nevertheless, the number of intellectual property lawyers in the USA is growing faster than the amount of research (Barton, 2000).

In contrast, the International Seed Federation (2002) strongly supports the EDV concept favoring thresholds from approx. GD=0.20 to 0.25, as suggested by Smith and Smith (1989), because it allows taking the above mentioned new technological developments into account. Furthermore, a strict threshold would support classical "creative" plant breeding and prevent from plagiarized "cosmetic" breeding, without hindering additive improving plant breeding. It would also enable building up a legal basis for balanced agreements among breeders as well as between breeders and inventors of patented procedures or products.

We have shown that Type I (α) and Type II (β) errors of a given GD threshold T were dependent on the crop, the degree of polymorphism of the marker system within the particular germplasm pool, the set of markers used, and the applied distance measure (Heckenberger et al., 2004). For example, T=0.25 based on Rogers' distance resulted in a fairly low α=0.07 to detect a BC_1-derived dent line as an EDV and would, therefore, be a possible EDV threshold to discriminate F_2- and BC_1-derived dent lines. For flint lines, however, T=0.25 yielded a considerably higher α=0.18. A possible threshold to distinguish between BC_1-, and BC_2-derived flint lines would be T=0.10. For dent and introgression lines, α-values were smaller than 0.02 and β_T lower than 20%. This indicates that a threshold of T=0.10 would be too conservative to distinguish between BC_1- and BC_2-derived dent or introgression lines and would consequently state the development of a BC_2 to a protected variety as an accepted breeding procedure (Troyer and Rocheford, 2002).

In detail, ASSINSEL and SEPROMA proposed a two-stage threshold for the detection of EDVs with a "red zone" of $GD_{(P1,O)}<0.10$, where a variety should be judged as an EDV, a "green zone" of $GD_{(P1,O)}>0.15$ (SEPROMA) or $GD_{(P1,O)}>0.20$ (ASSINSEL), where a variety should be judged as an IDV, and an "orange zone" between the two particular thresholds, where additional information is necessary to decide whether a variety is essentially derived or not. For the proposal of ASSINSEL, this would indicate that breeding an F_2-derived progeny from a protected line would be an accepted breeding procedure, but 72%, 39%, and 19% of flint, dent, and introgression BC_1-progenies would fall into the

"orange zone". The proposal of SEPROMA would indicate that breeding a BC_1-derived progeny from a protected line would be accepted with 36%, 13%, and 7% of flint, dent, and introgression BC_2-progenies situated in the "orange zone". For the "red zone" of ASSINSEL and SEPROMA, the same conclusions hold true as stated above for T=0.10.

Threshold Scenarios already applied on other Crops

In contrast to the presented scenarios based on the construction of frequency distributions of GD for each particular level of relatedness, the breeders of lettuce (*Lactuca sativa* L.) adopted a different scenario for essential derivation (International Seed Federation, 2003) based on a reference set that represent the total of all protected lettuce varieties. They agreed that a variety is deemed to beessentially derived if its GD to the initial variety was smaller than 95% of all the pairwise GDs of the reference set, independently of the marker system or the marker set used. For a recommended standard set of AFLP primer combinations, this threshold amounts currently to GD=0.05 based on 1- Jaccard's (Jaccard, 1908) similarity coefficient.

In ryegrass (*Lolium perenne* L.), a provisional threshold of a squared Euclidean Distance of seven was adopted in 1999 with the intention of a critical review after five years. If the distance was seven or lower, the breeder of the IV may ask for ISF arbitration. The breeder of the putative EDV will have to show that he has not practiced essential derivation from the IV. The arbitrators also have the right to ensure that the putative IV is not itself an EDV from a preexisting variety (Roldan Ruiz et al., 2000a).

All above mentioned threshold scenarios, including the scenarios developed in this study, depend more or less on the choice of reference sets of varieties or inbred lines to adjust the thresholds according to a crop or germplasm pool. The choice of varieties to be included into reference sets for the development of thresholds is, therefore, a crucial issue for the identification of EDVs. As genetic diversity within a certain crop may differ between countries or growing regions, a creation of the reference sets, representative for the crop or germplasm pool, may lead to problems. Consequently, thresholds should also be specific for the region they are developed for and the set of reference varieties must be assembled with caution.

Influence of intra-varietal Variation and Lab Error on the EDV Concept

A considerable variation between accessions of the same maize line, caused by lab error, PCR artefacts, or heterogeneity within varieties due to mutation or outcrossing, was observed for both SSRs and AFLPs (Heckenberger et al., 2002; 2003). Our results for SSRs confirmed a study of Gethi et al. (2002), who reported a variation in SSRs of approx. 8% between sources of the same inbred line. In addition, Vigouroux et al. (2002) and Matsuoka et al. (2002) reported considerable mutation rates, particularly for SSRs with a di-repeat motif, that were higher than expected for the natural mutation rate of genomic DNA.

Within parent-progeny triplets, this intra-varietal variation leads primarily to non-parental alleles (NPAs) in progeny lines. For SSRs, NPAs were found for 4.2% of all progeny data points, which was considerably lower than reported by Bernardo and Kahler (2001). The size differences from the corresponding parental alleles ranged from 1 to 81bp with a mean of 14bp, but a considerable portion of NPAs differed only 1-3bp from their corresponding parental alleles and could, therefore, be re-scored and assigned to their parental alleles. NPAs were detected in 2.2% of all AFLP progeny data points, whereas 45% of all AFLP markers showed an NPA in at least one triplet. In addition, the number of NPAs per triplet was highly correlated between SSRs and AFLPs.

The occurrence of NPAs decreases the correlation between the marker-estimated GD and the true GD, and should, therefore,be avoided as much as possible. Non-parental alleles were observed for a higher percentage of SSRs than of AFLPs, due to the lower error rate of a dominant marker system such as AFLPs (Heckenberger et al., 2003). In addition, the frequency distribution of size differences between NPAs and their corresponding parental alleles indicates that NPAs for SSRs were mainly caused by artificial stutter bands (Smith et al., 1997) or 1bp-differences between a parental and a progeny allele (Heckenberger et al., 2002). Hence, we recommend to avoid the use of SSRs with di-nucleotide repeat motifs for identification of EDVs, to minimize the probability of the occurrence of stutter bands and to reduce the risk of mutations (Vigouroux et al., 2002). However, the influence of NPAs on the selectivity of a particular marker system was rather small, and could be neglected after a cautious re-scoring of data.

For closely related genotypes, intra-varietal variation generally leads to an over-estimation of GDs, to the benefit of the breeder of the putative EDV. The breeders of the IVs should, therefore, warrant a high level of homogeneity in their inbred lines for their own benefits. Consequently, we strongly recommend increased levels of homogeneity of maize inbred lines before applying for plant varietal protection, as well as replications of lab assays to minimize experimental errors.

The use of Molecular Markers for DUS Testing

In several crops, *e.g.*, oilseed rape (*Brassica napus* L.) (Lombard et al., 2000), soybean (*Glycine max* L.) (Giancola et al., 2002), and maize (Dillmann et al., 1997), the use of molecular markers for testing of distinctness was evaluated. A common result of all the above-mentioned studies, including the present thesis, was that cultivars indistinguishable by morphological descriptors differed considerably in their banding patterns revealed by molecular markers.

This indicates that molecular markers offer the possibility for a more accurate comparison of varieties than morphological traits do. These comparisons, however, might be too accurate to observe distinctness on the basis of single bands because of the limited reproducibility of molecular marker data due to PCR artifacts. Consequently, observing distinctness on the basis of molecular markers would require certain thresholds for distinctness, similar to EDV thresholds to observe on conformity. All the above-mentioned authors suggested, therefore, the use of phenotypic characters for DUS testing with only an additional application of molecular markers.

Factors influencing the Relationship between *f* and GD

The coancestry coefficient (*f*) (Malécot, 1948) between parental lines and progenies was used in this study as a benchmark for breeding procedures applied for breeding the in the derivation of progenies. Factors influencing the relationship between *f* and GD were considered important for the validation of the ability of GDs to identify EDVs.

The value of *f* is defined as the probability that two homologous genes taken at random, one from each individual, are identical by descent (*ibd*), *i.e.* they are copies of the same gene from a common ancestor. In contrast, the Genetic Similarity (1-GD) of lines is based on bands alike in state (*ais*), *i.e.* bands indistinguishable whether they are identical due to a common ancestor or due to the genetic background of the particular germplasm. Bands that were only *ais* but not *ibd*, subsequently designated as *oais*, were ignored in calculating *f*, but remain considered for the calculation of GDs as they were indistinguishable from genes that were *ibd*. Consequently, for a close relationship between GD and *f*, the fraction of bands *oais* should be small (Messmer et al., 1993).

As only GDs based on bands *ais* can currently be calculated, we used $\hat{\mu}_{GD_{(P1,P2)}}$ of unrelated lines (*f*=0) within the same material group or germplasm pool as an estimator of the proportion of bands being *oais*. A method that might unravel the proportion of bands *ibd* and the conditional proportion of bands *oais* was proposed by Bernardo et al. (1996) using an iterative approach on the basis of known parent-progeny relationships. Nevertheless, the estimation of identity by descent with poorly or unknown pedigree relationships, as in the case of EDV, remains an unsolved problem.

In addition, *f* is based on several simplifying but mostly unrealistic assumptions (Melchinger et al., 1991). The first assumption (all lines in the pedigree pathway are homogeneous and homozygous) may be justified for most of the highly inbred lines in this study, but may not be true for all lines used. For some lines examined in this study, up to 25% heterozygous SSR loci were detected, although they were highly inbred. Violation of the second assumption (lines with no common parentage have *f*=0) leads to an underestimation of *f* if progenitors are, in fact, related. The third assumption (lines derived from a cross obtained half of the genome from each parent) is most disputable, as observed in the present study.

Furthermore, the relationship between *f* and GD is affected by selection, drift, and mutation. As the genetic model used for the simulation study was allowing for drift, but not for either selection or mutation, the good fit of observed and simulated data for F_2-derived progenies indicates, that the variation in *p* for F_2-derived lines was mainly caused by genetic drift, whereas the influences of selection or mutation on *p* were negligible. This result was in agreement with the results published by Bernardo and Kahler (2001), who reported

that the mean parental contribution for unselected F_2-progenies was close to the expected value of 0.50. Moreover, they found that the selection of progeny lines tended to increase the frequency of alleles of the parent selected for, whereas no significant differences with unselected progenies were observed.

The Use of Computer Simulations

The present triplet studies were carried out with empirical distributions for GD values between parents and progenies ($GD_{(P1,O)}$). Due to a limited number of F_2-, BC_1-, and BC_2-derived progenies, an analytical description of the distribution of $GD_{(P1,O)}$ is not yet available. Simulations, as evaluated by Bohn et al. (2004) in a companion study, were, therefore, used as an alternative approach to derive the distribution of the test statistic. The simulations were conducted using the PLABSIM software package (Frisch et al., 2000). This software enables to flexibly alter different crop or genome specific parameters, such as chromosome length, marker density, or degree of polymorphism of applied markers. Their effect on the distribution of $GD_{(P1,O)}$ can, thus, be directly assessed.

Simulated $GD_{(P1,O)}$ values were calculated on the basis of mean and variance of GD values between unrelated lines of a particular germplasm pool. Depending on the accuracy of the estimation of $\mu_{GD_{(P1,P2)}}$ and $\sigma^2_{GD_{(P1,P2)}}$, the simulations proved to be a powerful tool to verify empirical distributions of $GD_{(P1,O)}$. Moreover, the simulations can be applied to simulate EDV scenarios for any diploid crop, if parameters $\mu_{GD_{(P1,P2)}}$ and $\sigma^2_{GD_{(P1,P2)}}$, as well as the number of chromosomes and the chromosome length are known accurately, even when no empirical data of $GD_{(P1,O)}$ is available.

Direct *vs.* indirect Measures of Conformity

Additionally to the use of molecular markers, which was proposed for identification of EDVs by various authors (Bernardo and Kahler, 2001; Dillmann et al., 1997; Roldan Ruiz et al., 2000a; Smith and Smith, 1989), the use of phenotypic descriptors, such as morphological traits or heterosis, is still under consideration (ASSINSEL, 2000; Roldan Ruiz et al., 2000b; Gilliland et al., 2000; International Seed Federation, 2002; Smith and Smith,

1989). Supporters of phenotypic data application claim that the term 'conform in the expression of its essential characteristics that result from the genotype' (UPOV, 1991) implies the use of phenotypic data rather than molecular data.

Numerous studies showed a triangular relationship between GDs based on molecular markers and morphological distances (MDs) based on phenotypic traits (Burstin and Charcosset, 1997; Dillmann and Guérin, 1998; Rebourg et al., 2001; Roldan Ruiz et al., 2000b). By contrast, several studies reported linear relationships and high correlations between GDs and mid-parent heterosis (Melchinger, 1999; Boppenmaier et al., 1993; Smith and Smith, 1989; Ajmone Marsan et al., 1998).

Opponents of the use of phenotypic data state that even highly heritable phenotypic traits can offer only a rough estimate of the true relatedness of two cultivars. Based on our results, GDs based on molecular markers have clear advantages for the identification of EDVs. First, molecular data provide a direct estimate of the true relatedness of two genotypes because they are unbiased by the environmental effects. Second, molecular data reflect the percentage of the genome in common between the IV and a putative EDV, whereas certain morphological traits may differ in their expression within different environments, thus requiring extensive field trials. Third, a large number of markers is available for genotyping cultivars of all crops, but only a limited number of morphological traits can be observed with reasonable financial and labor efforts. Forth, only a small part of the genome might be involved in the expression of morphological traits, whereas markers can be chosen explicitly to ensure an equal and dense coverage of the genome. Fifth, scoring of marker bands can be automated to a large extent (Ziegle et al., 1992), thus being objective and reproducible, whereas morphological data may vary due to the subjectivity of the scoring person(s) (Nuel et al., 2001). Having all those issues in mind, we suggest a redefinition of the term "essential characteristics" in the sense that marker bands can also be regarded as essential characteristics in the terms of the UPOV convention (UPOV, 1991).

Accepted *vs.* non-accepted Breeding Procedures

No agreement on accepted or non-accepted breeding procedures has been achieved so far in maize. Studies on the influence of somaclonal variation during transformation yielded high similarities between transformed maize lines and their isogenic non-transformed counterparts (Marhic et al., 1998; Murigneux et al., 1993). This indicates that transformed varieties will most likely be judged as EDVs (Borgo et al., 2002) from their isogenic counterparts, even if they are distinct from them.

Regarding the number of acceptable backcrosses, proponents of low GD thresholds (*e.g.,* GD≤0.15) state that the original UPOV convention gives the term "backrosses" in its plural form in the examples of breeding procedures yielding EDVs (ASSINSEL, 1999), indicating that at least one backcross to a protected variety should be accepted. In contrast, opponents of low GD thresholds argue that by developing a BC_1, up to 95% of the genome of the recurrent parent can be maintained by marker assisted selection, which is against the intention of using a variety as a source of initial variation. However, no consensus has currently been achieved.

Conclusion and Outlook

The present thesis provides the first detailed comparison of various distance measures on their ability to identify EDVs in maize. We have shown that for various reasons GDs based on molecular markers are superior to MDs or heterosis in reflecting the true genetic relationships of two cultivars. Consequently, procedures for the identification of EDVs should be developed with an emphasis on molecular marker technologies, rather than on phenotypic traits.

For future prospects, a growing number of markers will be available for each marker system, ensuring an increased precision of GD estimates (Foulley and Hill, 1999) and, therefore, reducing the probabilities of being judged for essential derivation by chance. In addition, new marker systems and techniques, such as single nucleotide polymorphisms (SNPs), or expression profiles in combination with microarrays or DNA chips will further increase the accuracy of molecular methods in estimating the true genetic relatedness of two cultivars. Finally, an adapted form of forensic approaches (Gill et al.,

1995; Graham et al., 2000), already applied successfully in human genetics for verification of parentage or disproving suspects, could aid in the identification of EDVs.

References

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

ASSINSEL. 1999. Essential derivation and dependence. Practical Information.

ASSINSEL. 2000. DUS testing: phenotype *vs.* genotype. ASSINSEL, Position paper adopted at the Rome Congress in May 2000.

Barton, J. H. 2000. Intellectual property rights: reforming the patent system. Science. 287:1933-1934.

Bernardo, R. and A. L. Kahler. 2001. North American study on essential derivation in maize: inbreds developed without and with selection from F_2 populations. Theor. Appl. Genet. 102:986-992.

Bernardo R, Murigneux A, and Karaman Z. 1996. Marker-based estimates of identity by descent and alikeness in state among maize inbreds. Theor. Appl. Genet. 93:262-267.

Bohn, M., M. Frisch, and A. E. Melchinger. 2004. A statistical framework for identification of essentially derived varieties obtained from biparental crosses of homozygous lines. I. Statistical theory. In Preparation.

Boppenmaier, J., A. E. Melchinger, G. Seitz, H. H. Geiger, and R. G. Herrmann. 1993. Genetic diversity for RFLPs in European maize inbreds. III. Performance of crosses within versus between heterotic groups for grain traits. Plant Breed. 111:217-226.

Borgo, L., Serani B., and S. Conti. 2002. Molecular profiling of genetically modified maize lines as related to the issue of essential derivation. Poster presented at the Annual Congress of the Italian Society of Agricultural Genetics, Sept 18-21 2002, Girardini Naxos, Italy.

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

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

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

Duvick, D.N. 1984. Genetic contributions to yield gains of USA hybrid maize, 1930 to 1980. p. 15–47. *In* W.R. Fehr (ed.) Genetic contributions to yield gains of five major crop plants. CSSA Spec. Publ. 7. CSSA, Madison, WI.

Foulley, J. L. and W. G. Hill. 1999. On the precision of estimation of genetic distance. Genet. Sel. Evol. 31:457-464.

Frisch, M., M. Bohn, and A. E. Melchinger. 2000. Plabsim: Software for simulation of marker-assisted backcrossing. J. Hered. 91:86-87.

Gethi, J. G., J. A. Labate, K. R. Lamkey, M. E. Smith, and S. Kresovich. 2002. Ssr variation in important us maize inbred lines. Crop Sci. 49:951-957.

Giancola, S., S. Marcucci Poltri, P. Lacaze, and H. E. Hopp. 2002. Feasibility of integration of molecular markers and morphological descriptors in a real case study of a plant variety protection system for soybean. Euphytica. 127:95-113.

Gill, P., C. P. Kimpton, A. Urquhart, N. Oldroyd, E. S. Millican, S. K. Watson, and T. J. Downes. 1995. Automated short tandem repeat (STR) analysis in forensic casework: a strategy for the future. Electrophoresis. 16:1543-1552.

Gilliland, T. J., R. Coll, E. Calsyn, M. de. Loose, M. J. T. v. Eijk, and I. Roldan Ruiz. 2000. Estimating genetic conformity between related ryegrass (*Lolium*) varieties. 1. Morphology and biochemical characterisation. Mol-breed. 6:569-580.

Graham, J., J. Curran, and B. S. Weir. 2000. Conditional Genotypic Probabilities for Microsatellite Loci. Genetics. 155:1973-1980.

Heckenberger, M., M. Bohn, and A. E. Melchinger. 2004. A statistical framework for identification of essentially derived varieties obtained from biparental crosses of homozygous lines. II. SSR data from maize. Crop Sci. submitted.

Heckenberger, M., M. Bohn, J. S. Ziegle, L. K. Joe, J. D. Hauser, M. Hutton, and A. E. Melchinger. 2002. Variation of DNA fingerprints among accessions within maize inbred lines and implications for identification of essentially derived varieties. I. Genetic and technical sources of variation in SSR data. Mol. Breed. 10 :181-191.

Heckenberger, M., J. Rouppe van der Voort, A. E. Melchinger, J. Peleman, and M. Bohn. 2003. Variation of DNA fingerprints among accessions within maize inbred lines and implications for identification of essentially derived varieties. II. Genetic and technical sources of variation in AFLP data and comparison to SSR data. Mol. Breed. 12:97-106.

International Seed Federation. 2002. ISF view on Intellectual Property.

International Seed Federation. 2003. 2002/2003 Progress Report of the Secretariat.

Jaccard, P. 1908. Nouvelles recherches sur la distribution florale. Bull. Soc. Vaud. Sci. Nat. 44:223-270.

Lombard, V., C. P. Baril, P. Dubreuil, F. Blouet, and D. Zhang. 2000. Genetic relationships and fingerprinting of rapeseed cultivars by AFLP: consequences for varietal registration. Crop Sci. 40:1417-1425.

Malécot, G. 1948. Les mathematiques de l'heredité. p. Masson & Cies, Paris.

Marhic, A., S. Anatoine-Michard, J. Bordes, M. Pollacsek, A. Marigneux, and M. Beckert. 1998. Genetic improvement of anther culture response in maize: relationships with molecular, mendelian and agronomic traits. Theor. Appl. Genet. 97:520-525.

Matsuoka, Y., S. E. Mitchell, S. Kresovich, M. Goodman, and J. Doebley. 2002. Microsatellites in Zea - Variability, Patterns of Mutations, and Use for Evolutionary Studies. Theor. Appl. Genet. 104 :436-450.

Melchinger, A.E. 1999. Genetic Diversity and Heterosis. International Symposium on Genetics and Exploitation of Heterosis in Crop Plants. Mexico City. p. 99-118. *In* J.G. Coors and S. Pandey (ed.) The genetics and exploitation of heterosis in crops. ASA, CSSA, and SSSA, Madison, WI.

Melchinger, A. E., M. M. Messmer, M. Lee, W. L. Woodman, and K. R. Lamkey. 1991. Diversity and relationships among U.S. maize inbreds revealed by restriction fragment length polymorphism. Crop Sci. 31:669-678.

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

Multani, D. S., S. P. Briggs, M. A. Chamberlin, J. J. Blakeslee, A. S. Murphy, and G. S. Johal. 2003. Loss of an MDR transporter in compact stalks of maize *br2* and sorghum *dw3* mutants. Science. 302:81-84.

Murigneux, A., D. Barloy, P. Leroy, and M. Beckert. 1993. Molecular and morphological evaluation of doubled haploid lines in maize. 1. Homogeneity within DH lines. Theor. Appl. Genet. 86:837-842.

Nuel, G., S. Robin, and Baril C.P. 2001. Predicting distanced using a linear model: the case of varietal distinctness. J. Appl. Stat. 28:607-621.

Rebourg, C., B. Gouesnard, and A. Charcosset . 2001. Large scale molecular analysis of traditional European maize populations. Relationships with morphological variation. Heredity. 86:574-587.

Roldan Ruiz, I., E. Calsyn, T. J. Gilliland, R. Coll, M. J. T. van Eijk, and De LooseM. 2000a. Estimating genetic conformity between related ryegrass (*Lolium*) varieties. 2. AFLP characterization. Mol. Breed. 6:593-602.

Roldan Ruiz, I., F. A. v. Eeuwijk, T. J. Gilliland, P. Dubreuil, C. Dillmann, J. Lallemand, M. de. Loose, and C. P. Baril. 2000b. A comparative study of molecular and morphological methods of describing relationships between perennial ryegrass (*Lolium perenne* L.) varieties. Theor. Appl. Genet. 103:1138-1150.

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.

Smith, J. S. C. and O. S. Smith. 1989. The description and assessment of distances between inbred lines of maize: II. The utility of morphological, biochemical, and genetic descriptors and a scheme for testing of distinctiveness between inbred lines. Maydica. 34:151-161.

The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature. 408:796-815.

Thornsberry, J. M., M. M. Goodman, J. Doebley, S. Kresovich, D. Nielsen, and E. S. Buckler. 2001. *Dwarf8* polymorphisms associate with variation in flowering time. Nat. Genet. 28:286-289.

Troyer, A. F. and T. R. Rocheford. 2002. Germplasm ownership: related corn inbreds. Crop Sci. 42:3-11.

UPOV. 1991. International convention for the protection of new varieties of plants.

Vigouroux, Y., J. S. Jaqueth, Y. Matsuoka, O. S. Smith, W. D. Beavis, J. S. C. Smith, and J. Doebley. 2002. Rate and pattern of mutation at microsatellite loci in maize. Mol. Biol. Evol. 19:1251-1260.

Walsh, B. 2000. Quantitative genetics, genomics, and the future of plant breeding. University of Arizona.

Walsh, B. 2001. Quantitative genetics in the age of genomics. Theor. Pop. Biol. 59:175-184.

Ziegle, J. S., K. P. Corcoran, L. Nie, P. E. Mayrand, L. B. Hoff, L. J. McBride, M. N. Kronick, and S. R. Diehl. 1992. Application of automated DNA sizing technology for genotyping microsatellite loci. Genomics. 14:1026-1031.

7. Summary

The ‘breeder’s exemption’ as fixed in the UPOV convention on plant variety protection allows the use of protected germplasm for the development of new plant varieties. The aim of this concept is the creation of new genetic variation to guarantee a continuous breeding progress. However, the use of molecular markers in backcrossing programs and genetic engineering has created the technical basis to develop new plant varieties without original breeding efforts. Therefore, the concept of ‘essential derivation’ was implemented into the 1991 Act of the UPOV convention to distinguish between varieties that resulted from intensive and creative selection programs and cultivars that were developed without major genetic changes from these former varieties. Accordingly, a variety is deemed to be essentially derived from an initial variety (IV), if it (i) was predominantly derived from the IV, (ii) is clearly distinguishable from the IV, and (iii) genetically conforms to the IV in the expression of it’s essential characteristics.

The goal of this thesis was to evaluate and compare different approaches to assess conformity in the expression of the essential characteristics between IV and essentially derived varieties (EDVs) and to derive a theoretical and experimental basis for the development of thresholds to distinguish between independently derived varieties and EDVs in maize (*Zea mays* L.). The main focus was set on the evaluation of genetic distances based on ‘simple sequence repeats’ (SSRs) and ‘amplified fragment length polymorphisms’ (AFLPs) as well as the factors contributing to the GD between parental inbreds and their progeny lines. Furthermore, the ability of heterosis and morphological distances for identification of EDVs was examined. In detail, the objectives were to (1) investigate genetic and technical sources of variation in data derived from simple sequence repeats (SSR) and amplified fragment length polymorphism (AFLPs) within maize inbreds and assess their impact on identification of EDVs, (2) analyze the factors influencing genetic distances (GD) based on SSRs and AFLPs between related maize inbred lines, (3) investigate the power of SSR- and AFLP-based GD estimates, morphological distances and heterosis for discriminating between progenies derived from F_2, BC_1, and BC_2 populations, (4) exemplify theoretical and simulated results with experimental data, and (5) draw conclusions with regard to EDV thresholds suggested in the literature.

A total of 220 flint, dent, and US maize inbred lines was genotyped with 100 SSRs equally distributed across the maize genome. The 220 lines comprised 163 triplets. A triplet consisted of one progeny and both parental lines, where the former was developed from an F_2-, BC_1-, or BC_2 population. A subset of 58 lines (38 triplets) was genotyped additionally with 20 AFLP primer combinations. Furthermore, morphological traits and heterosis were observed for these 38 triplets in a field experiment over two years and three locations.

In addition, two to five accessions from nine inbred lines and five doubled haploid (DH) lines were taken from different sources or drawn as independent samples from the same seed lot and genotyped with SSRs and AFLPs to examine the variation of SSR and AFLP data within maize inbred lines. The GD between accessions of the same inbred or DH line amounted to 0.03 for SSRs and to 0.01 for AFLPs and was, therefore, of minor importance for identification of EDVs.

Parental genome contribution to F_2-derived lines estimated with SSRs ranged from 0.25 to 0.70 with a mean of 0.49. Deviations of the mean from the expected value can be explained by the occurrence of non-parental bands that were detected in 4% of all datapoints. The parental contribution of the recurrent parent to BC_1-derived progenies varied from 0.44 to 0.79 with an average of 0.64 and was significantly smaller than the expected parental genome contribution of 0.75. The distributions of GD values for parental lines and their F_2- and BC_1-derived progeny overlapped for simulated as well as for experimental data.

An analysis of variance revealed that for F_2-derived progeny lines 34% of the variance of the GD between parent and progeny line ($\sigma^2_{GD_{(P1;O)}}$) were explained by the variance of the GD between the parental lines ($\sigma^2_{GD_{(P1;P2)}}$) and 66% by the variance of the parental contribution (σ^2_p). For BC_1-derived progenies, $\sigma^2_{GD_{(P1;O)}}$ was largely independent of $\sigma^2_{GD_{(P1;P2)}}$, as more than 95% of $\sigma^2_{GD_{(P1;O)}}$ were explained by σ^2_p.

Assuming that the derivation of a line from an F_2 population was an accepted breeding procedure and the derivation from a BC_1 population would not be accepted, we observed Type II errors (β) ranging from 0.23 to 0.37 depending on the germplasm pool for a given Type I error (α) of 0.05. For a threshold between BC_1 and BC_2, β ranged from 0.40 to 0.60 with an increasing tendency for higher BC levels. For fixed GD thresholds of

T=0.25, 0.20, 0.15, and 0.10 suggested in the literature, substantial differences for α and β were found between different germplasm pools. Therefore, thresholds need to be gene pool specific and different thresholds for potential EDVs from intra-pool crosses than for progenies from inter-pool crosses must be applied.

Discrimination of F_2-, BC_1-, and BC_2-derived progeny lines on the basis of heterosis and morphological distances revealed β values ranging from 0.50 to 0.95 depending on the trait or combination of traits. Therefore, heterosis and morphological distances were fairly inappropriate tools for identification of EDVs due to the larger overlaps of F_2-, BC_1-, and BC_2-distributions compared to GDs based on molecular markers.

In general, SSRs and AFLPs were the most adequate tools to uncover close pedigree relationships between maize inbred lines and to discriminate among lines derived with accepted or non-accepted breeding procedures. Therefore, the results presented in this study provide an example for identification of EDVs and can be transferred to other diploid crops by adjusting the corresponding thresholds.

8. Zusammenfassung

Das in der UPOV-Konvention festgelegte „Züchterprivileg“ sichert allen Züchtern die freie Verwendung von geschützten Sorten zur Schaffung neuer genetischer Variation und ist damit Garant für nachhaltigen Zuchtfortschritt. Da das Züchterprivileg jedoch auf verschiedene Weise unterlaufen werden kann, z. B. bei Re-Selektion innerhalb von geschützten Sorten, fortgesetzter Rückkreuzung zu geschützten Sorten, oder durch punktuelle gentechnische Veränderungen geschützter Sorten, wurde von der UPOV 1991 das Konzept der „abgeleiteten Sorten“ (essentially derived varieties, EDVs) eingeführt, um diesen Missbrauch zu verhindern. Eine Sorte ist demnach von einer Ausgangssorte abgeleitet, wenn (i) die Sorte sich zumindest in einem Registermerkmal von der Ausgangssorte unterscheidet, (ii) der Züchter die Sorte hauptsächlich aus der Ausgangssorte entwickelt hat und (iii) die Sorte weitestgehend genetisch identisch ist mit der Ausgangssorte.

Hauptziel dieser Arbeit war es, bei Mais verschiedene Methoden zur Feststellung der genetischen Übereinstimmung zwischen Ausgangssorte und potentiellen EDVs zu untersuchen und miteinander zu vergleichen, um eine theoretische und experimentelle Basis zur Festlegung von Grenzwerten zur Unterscheidung von EDVs und unabhängigen Sorten zu erarbeiten. Im Vordergrund stand dabei die Evaluierung von genetischen Distanzen (GDs) basierend auf molekularen Markern, wie „simple sequence repeats“ (SSRs) und „amplified fragment length polymorphisms“ (AFLPs) sowie Faktoren, welche die GD zwischen Eltern und Nachkommen bestimmen. Ebenso wurde die Möglichkeit untersucht, morphologische Merkmale oder Heterosis zur Bestimmung von EDVs zu verwenden. Einzelziele der Arbeit waren: (1) die Untersuchung der Variation von SSRs und AFLPs innerhalb von Maisinzuchtlinien; (2) die Bestimmung der Variation des Genombeitrags von Elternlinien zu ihren Nachkommenlinien; (3) die Unterscheidung von Maisinzuchtlinien aus F_2-, BC_1- und BC_2-Nachkommenschaften anhand von SSR- und AFLP- basierten GDs sowie morphologischen Merkmalen und Heterosis; (4) die Verifikation von theoretischen und simulierten Daten aus einer Begleitstudie und (5) die Bewertung der Ergebnisse im Hinblick auf verschiedene in der Literatur vorgeschlagene EDV-Grenzwerte.

Insgesamt wurden 220 Flint-, Dent- und US-Maisinzuchtlinien mit 100 gleichmäßig über das Maisgenom verteilten SSRs genotypisiert. Darunter befanden sich 163

Triplets bestehend aus einer Nachkommenlinie und beiden Elternlinien, wobei erstere entweder aus einer F_2-, BC_1-, oder BC_2- Population entwickelt wurde. Ein Teilsatz aus 58 Linien (38 Triplets) wurde zusätzlich mit 20 AFLP-Primerkombinationen genotypisiert. In einem dreiortigen Feldexperiment über zwei Jahre wurden darüber hinaus morphologische Merkmale und Heterosis für diese 38 Triplets erfasst.

Eine SSR- und AFLP-Analyse verschiedener Akzessionen derselben Maisinzuchtlinie, beispielsweise aus der Erhaltungszüchtung verschiedener Züchter oder aus unterschiedlichen Vermehrungsstufen, wurde zur Untersuchung der Variation von Markerdaten innerhalb von Maisinzuchtlinien herangezogen. Die GD zwischen Akzessionen derselben Linie betrug dabei im Mittel 0,03 für SSRs bzw. 0,01 für AFLPs und erwies sich damit als vernachlässigbar für die Bestimmung von EDV-Grenzwerten.

Der mit SSRs bestimmte elterliche Genombeitrag zu den F_2-abgeleiteten Nachkommenlinien variierte zwischen 25% und 70% bei einem Mittel von 49%. Die Abweichung des Mittelwertes vom Erwartungswert von 50% lassen sich mit dem Auftreten nichtelterlicher Banden erklären, die bei 4% aller Datenpunkte detektiert wurden. Der Genombeitrag des rekurrenten Elters zu den BC_1-abgeleiteten Nachkommenlinien variierte von 44% bis 79% bei einem Mittel von 64% und war dabei signifikant (P<0,05) kleiner als der Erwartungswert von 75%. Sowohl bei simulierten als auch bei experimentellen Daten wurden Überlappungen zwischen den GD-Verteilungen von F_2- und BC_1- abgeleiteten Linien zu ihren Eltern gefunden.

Eine Varianzanalyse ergab, dass bei F_2-abgeleiteten Nachkommenlinien die Varianz der GD zwischen Elter und Nachkommenlinie ($\sigma^2_{GD_{(P1;O)}}$) zu ca. 34% durch die Varianz der GD der Elternlinien ($\sigma^2_{GD_{(P1;P2)}}$) und zu ca. 66% durch die Varianz des elterlichen Genombeitrags (σ^2_p) erklärt wird. Für BC_1-abgeleitete Nachkommenlinien erwies sich $\sigma^2_{GD_{(P1;O)}}$ als weitestgehend unabhängig von $\sigma^2_{GD_{(P1;P2)}}$, da über 95% des Wertes von $\sigma^2_{GD_{(P1;O)}}$ durch σ^2_p erklärt wurden.

Unter der Annahme, dass die Ableitung einer Linie aus einer F_2-Population ein akzeptiertes Verfahren wäre, die Ableitung aus einer BC_1-Population jedoch nicht akzeptabel, wurde bei einem angenommenen Fehler 1. Art (α) von 0,05 je nach Genpool und verwendetem Markersystem ein Fehler 2. Art (β) von 0,23 bis 0,37 festgestellt. Für einen

Grenzwert zwischen BC_1 und BC_2 liegt β zwischen 0,40 und 0,60 mit steigender Tendenz für höhere BC-Generationen. Für die in der Literatur vorgeschlagenen fixen Grenzwerte von T=0,25, 0,20, 0,15, und 0,10 wurden beträchtliche Differenzen für α und β zwischen verschiedenen Genpools gefunden, was die Anwendung genpoolspezifischer Grenzwerte nahe legt. Ebenso sollten spezifische Grenzwerte für potentielle EDVs aus intra-pool Kreuzungen gegenüber inter-pool Kreuzungen entwickelt werden.

Die Trennung von F_2-, BC_1- und BC_2-abgeleiteten Nachkommenlinien auf der Basis von morphologischen Distanzen und Heterosis ergab für α=0,05 β-Werte zwischen 0,50 und 0,95 je nach Merkmal bzw. Merkmalskombination. Sie erwies sich somit als wenig geeignet zur Identifikation von EDVs, da die F_2 -, BC_1- und BC_2-Verteilungen deutlich stärker überlappten als die mit Marker ermittelten GDs.

Generell erwiesen sich SSRs und AFLPs als am besten geeignet, enge Verwandtschaftsbeziehungen zwischen Maisinzuchtlinien aufzudecken und zwischen Linien zu unterscheiden, die mittels akzeptierten bzw. nicht akzeptierten Zuchtverfahren erstellt wurden. Die in dieser Studie vorgestellten Ergebnisse können somit als exemplarisch für die Identifikation von EDVs betrachtet werden und können durch Adjustierung der entsprechenden Schwellenwerte auf andere diploide Kulturarten übertragen werden.

9. Appendix

Table A. List of parents and progeny lines genotyped in the study of chapter 4.

Flint			**Dent**			**US**		
	parent	progeny		parent	progeny		parent	progeny
C	x		623	x		52BS26	x	
CD1		x	663	x		B100	x	
CD167	x		1720	x		B101	x	
CD2		x	1784	x		B102	x	
CD3		x	1790	x		B105-1	x	
CD4		x	2589	x		B106	x	
CE1		x	CO255	x		B97	x	
CE2		x	D06	x		B99	x	
CE3		x	D16	x	x	BS10-n-1	x	
CE4		x	D17	x	x	BS11-n-12	x	
CE5		x	D19	x		D5BS26	x	
CE6		x	D22	x		DBS26	x	
CE7		x	D23		x			
CO255	x		D24		x	**Introgression**		
D	x		D38		x	FB-1		x
D145	x		D408		x	FB-10		x
D146	x		D46		x	FB-11		x
D147	x		D48		x	FB-12		x
D149	x		D504		x	FB-13		x
D150	x		D60		x	FB-14		x
D167	x		D61		x	FB-15		x
D171	x		D62		x	FB-3		x
D503	x		D63		x	FB-6		x
DK105	x		D64		x	FB-8		x
E	x		D66		x	FB-9		x
F004		x	D67		x	L005		x
F006		x	D701		x	L007		x
F009		x	D704	x		L012		x
F011		x	D711		x	L017		x
F014		x	D726		x	L019		x
F017		x	D757	x		L020		x
F018	x	x	D83		x	L021		x
F020		x	DK105	x		P037		x
F023		x	F	x		S015		x
F027		x	FG1		x	S016		x

Table A (continued). List of parents and progeny lines genotyped in the study of chapter 4.

Flint			**Dent**			**Introgression**		
	parent	progeny		parent	progeny		parent	progeny
F036		x	FH1		x	S017		x
F038		x	G	x		S018		x
F039		x	H	x		S019		x
F040		x	LO977	x		S020		x
F042		x	M001		x	S021		x
F2U-1		x	M002		x	S022		x
F2U-10		x	M003		x	S023		x
F2U-11		x	M006		x	S024		x
F2U-12		x	M009		x	S025		x
F2U-13		x	N001	x		S026		x
F2U-14		x	NQ1		x	S027		x
F2U-15		x	NQ10		x	S028		x
F2U-16		x	NQ2		x	S030		x
F2U-2		x	NQ3		x	S031		x
F2U-3		x	NQ4		x	UB-1		x
F2U-4		x	NQ5		x	UB-10		x
F2U-5		x	NQ6		x	UB-11		x
F2U-7		x	NQ7		x	UB-12		x
F2U-8		x	NQ8		x	UB-2		x
F2U-9		x	NQ9		x	UB-3		x
F7D-1		x	P009	x	x	UB-6		x
F7D-2		x	P015		x	UB-7		x
F7D-3		x	P017		x	UB-8		x
F7D-4		x	P02425		x	UB-9		x
F7U-1		x	P041		x			
F7U-2		x	P042		x			
F7U-3		x	P043		x			
F7U-4		x	PD014-1		x			
FL025		x	PD014-2		x			
I	x		PD014-3		x			
ICD-2		x	Q001	x				
ICD-4		x	R0346	x				
ICD-5		x	R2188	x				
ICU-2		x	UH200	x	x			
ICU-3		x	UH201		x			
ICU-4		x	UH300		x			
ICU-5		x	D09	x				

Table A (continued). List of parents and progeny lines genotyped in the study of chapter 4.

Flint			**Dent**			**Introgression**		
	parent	progeny		parent	progeny		parent	progeny
ICU-6		x	D30	x				
ICU-7		x	D89	x				
ICU-8		x						
ICU-9		x						
J	x							
K	x							
K1077	x							
K2050B	x							
K2148C		x						
K3164B	x							
L	x							
R1143	x							
S208		x						
UH001	x	x						
UH002		x						
UH003		x						

Table B. Parent-offspring triplets of chapter 4 and their pedigree relationships.

Parent 1	Parent 2	Progeny	$f_{(P1,P2)}$	Type
Dent				
0663	1720	D60	0.000	F_2
0663	D16	D24	0.000	F_2
0663	D06	D19	0.000	F_2
2589	0663	D22	0.000	F_2
0663	D408	D83	0.000	F_2
D06	D61	P017	0.000	F_2
D83	D46	UH300	0.000	F_2
LO977	R0346	R2188	0.000	F_2
D704	D726	M003	0.000	F_2
F	G	FG1	0.000	F_2
F	H	FH1	0.000	F_2
1790	D06	D23	0.000	F_2
N001	Q001	NQ1	0.000	F_2
N001	Q001	NQ2	0.000	F_2
N001	Q001	NQ3	0.000	F_2
N001	Q001	NQ4	0.000	F_2
N001	Q001	NQ5	0.000	F_2
DK105	0663	D711	0.000	F_2
N001	Q001	NQ6	0.000	F_2
N001	Q001	NQ7	0.000	F_2
N001	Q001	NQ8	0.000	F_2
N001	Q001	NQ9	0.000	F_2
N001	Q001	NQ10	0.000	F_2
P02425	UH200	PD014-1	0.010	F_2
P02425	UH200	P041	0.010	F_2
P02425	UH200	P042	0.010	F_2
P02425	UH200	PD014-2	0.010	F_2
P02425	UH200	P043	0.010	F_2
P02425	UH200	PD014-3	0.010	F_2
D23	D46	UH300	0.029	F_2
D64	D06	P009	0.219	F_2
D64	D06	P015	0.219	F_2
D711	D704	M006	0.375	F_2
D711	D704	M009	0.375	F_2
D711	D701	M001	0.406	F_2
D711	D757	M002	0.406	F_2
D23	D06	UH200	0.500	F_2
D23	D06	UH201	0.500	F_2
0663	1720	D63	0.000	BC_1
0663	1720	D64	0.000	BC_1
0663	1720	D66	0.000	BC_1

Table B (continued). Parent-offspring triplets of chapter 4 and their pedigree relationships.

Parent 1	**Parent 2**	**Progeny**	$f_{(P1,P2)}$	**Type**
0663	1720	D67	0.000	BC_1
0663	D16	D48	0.000	BC_1
0663	D17	D61	0.000	BC_1
0663	1784	D38	0.000	BC_1
0663	0623	D46	0.000	BC_1
0663	D504	D701	0.000	BC_1
0663	D504	D757	0.000	BC_1
0663	CO255	D704	0.000	BC_1
0663	D17	D62	0.000	BC_2
Flint				
D146	C	F018	0.000	F_2
D167	C	F020	0.000	F_2
CO255	D503	R1143	0.000	F_2
DK105	K2050B	K2148C	0.000	F_2
K1077	S208	K3164B	0.000	F_2
C	D145	CD1	0.000	F_2
C	D145	CD2	0.000	F_2
C	D145	CD3	0.000	F_2
C	D145	CD4	0.000	F_2
C	E	CE1	0.000	F_2
C	E	CE2	0.000	F_2
C	E	CE3	0.000	F_2
C	E	CE4	0.000	F_2
C	E	CE5	0.000	F_2
C	E	CE6	0.000	F_2
C	E	CE7	0.000	F_2
I	C	F023	0.000	F_2
J	C	F020	0.000	F_2
K	L	F027	0.000	F_2
CD167	F011	FL025	0.056	F_2
CD167	UH001	F038	0.060	F_2
F023	UH001	F039	0.060	F_2
F023	UH001	ICU-2	0.060	F_2
F023	UH001	ICU-3	0.060	F_2
F023	UH001	ICU-4	0.060	F_2
F023	UH001	ICU-5	0.060	F_2
F023	UH001	ICU-6	0.060	F_2
F023	UH001	ICU-7	0.060	F_2
F023	UH001	ICU-8	0.060	F_2
F023	UH001	ICU-9	0.060	F_2
F023	D171	F040	0.074	F_2
F023	D171	ICD-2	0.074	F_2
F023	D171	F042	0.074	F_2
F023	D171	ICD-4	0.074	F_2

Table B (continued). Parent-offspring triplets of chapter 4 and their pedigree relationships.

Parent 1	Parent 2	Progeny	$f_{(P1,P2)}$	Type
F023	D171	ICD-5	0.074	F_2
D149	D150	F009	0.106	F_2
D149	D150	F011	0.106	F_2
D149	D150	UH002	0.106	F_2
D149	D171	UH003	0.128	F_2
D149	D171	F014	0.128	F_2
D149	D171	F017	0.128	F_2
D147	D171	F004	0.133	F_2
D147	D171	UH001	0.133	F_2
D147	D171	F006	0.133	F_2
F020	UH001	F2U-1	0.159	F_2
F020	UH001	F2U-2	0.159	F_2
F020	UH001	F2U-3	0.159	F_2
F020	UH001	F2U-4	0.159	F_2
F020	UH001	F2U-5	0.159	F_2
F020	UH001	F036	0.159	F_2
F020	UH001	F2U-7	0.159	F_2
F020	UH001	F2U-8	0.159	F_2
F020	UH001	F2U-9	0.159	F_2
F020	UH001	F2U-10	0.159	F_2
F020	UH001	F2U-11	0.159	F_2
F020	UH001	F2U-12	0.159	F_2
F020	UH001	F2U-13	0.159	F_2
F020	UH001	F2U-14	0.159	F_2
F020	UH001	F2U-15	0.159	F_2
F020	UH001	F2U-16	0.159	F_2
F027	UH001	F7U-1	0.171	F_2
F027	UH001	F7U-2	0.171	F_2
F027	UH001	F7U-3	0.171	F_2
F027	UH001	F7U-4	0.171	F_2
F027	D171	F7D-1	0.243	F_2
F027	D171	F7D-2	0.243	F_2
F027	D171	F7D-3	0.243	F_2
F027	D171	F7D-4	0.243	F_2
Introgression triplets				
UH300	B105-1	UB-7	0.000	F_2
UH300	B105-1	UB-8	0.000	F_2
UH300	B105-1	UB-9	0.000	F_2
UH300	B105-1	UB-10	0.000	F_2
UH300	B105-1	UB-11	0.000	F_2
D30	BS10-n-1	S025	0.000	F_2
D06	BS10-n-1	S026	0.000	F_2
D06	BS10-n-1	S027	0.000	F_2
D06	BS10-n-1	S028	0.000	F_2

Table B (continued). Parent-offspring triplets of chapter 4 and their pedigree relationships.

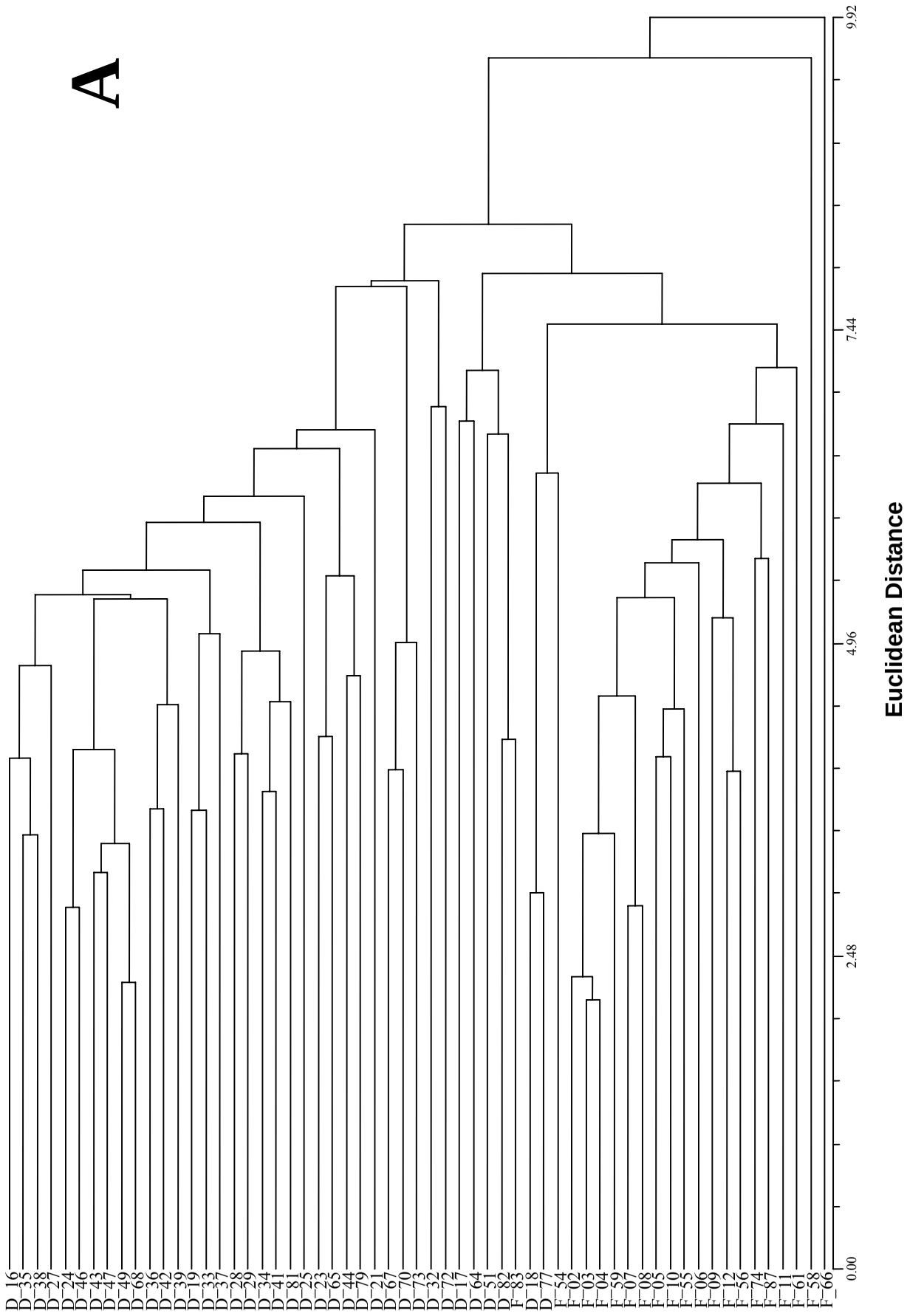
Parent 1	**Parent 2**	**Progeny**	$f_{(P1,P2)}$	**Type**
D5BS26	B99	L007	0.000	F_2
B105-1	UH200	UB-12	0.000	F_2
UH200	B105-1	UB-1	0.000	BC_1
UH200	B105-1	S031	0.000	BC_1
UH200	B105-1	UB-2	0.000	BC_1
UH300	B101	UB-3	0.000	BC_1
UH300	B101	P037	0.000	BC_1
UH300	B105-1	UB-6	0.000	BC_1
BS10-n-1	D23	S015	0.000	BC_1
BS10-n-1	D23	S016	0.000	BC_1
BS10-n-1	D23	S017	0.000	BC_1
BS11-n-12	D09	S018	0.000	BC_1
BS11-n-12	D09	S019	0.000	BC_1
BS11-n-12	D09	S020	0.000	BC_1
BS11-n-12	D89	S021	0.000	BC_1
BS11-n-12	D89	S022	0.000	BC_1
BS11-n-12	D89	S023	0.000	BC_1
BS11-n-12	D89	S024	0.000	BC_1
B105-1	UH200	S030	0.000	F_2
52BS26	B97	L005	0.000	F_2
DBS26	F028	L012	0.000	F_2
F011	B97	FB-1	0.000	F_2
F011	B97	L017	0.000	F_2
F011	B97	FB-3	0.000	F_2
F011	B97	L019	0.000	F_2
F011	B97	L020	0.000	F_2
F011	B97	FB-6	0.000	F_2
F011	B97	L021	0.000	F_2
F011	B100	FB-8	0.000	F_2
F011	B100	FB-9	0.000	F_2
F011	B100	FB-10	0.000	F_2
F011	B100	FB-11	0.000	F_2
F011	B100	FB-12	0.000	F_2
F011	B102	FB-13	0.000	F_2
F011	B102	FB-14	0.000	F_2
F011	B106	FB-15	0.000	F_2

Table C. Parent offspring triplets of chapter 5 and their pedigree relationships.

Parent 1	**Parent 2**	**Progeny**	$f_{(P1,P2)}$	**Type**
Flint				
D146	C1	F018	0.000	F_2
D167	C1	F020	0.000	F_2
CO255	D503	R1143	0.000	F_2
D149	D150	F009	0.106	F_2
D149	D150	F011	0.106	F_2
D149	D150	UH002	0.106	F_2
D149	D171	UH003	0.128	F_2
D149	D171	F014	0.128	F_2
D149	D171	F017	0.128	F_2
D147	D171	F004	0.133	F_2
D147	D171	UH001	0.133	F_2
D147	D171	F006	0.133	F_2
Dent				
0663	1720	D60	0.000	F_2
0663	D16	D24	0.000	F_2
0663	D06	D19	0.000	F_2
2589	0663	D22	0.000	F_2
0663	D408	D83	0.000	F_2
D06	D61	P017	0.000	F_2
D83	D46	UH300	0.000	F_2
D704	D726	M003	0.000	F_2
1790	D06	D23	0.000	F_2
D64	D06	P009	0.219	F_2
D64	D06	P015	0.219	F_2
D711	D704	M006	0.375	F_2
D711	D704	M009	0.375	F_2
D711	D701	M001	0.406	F_2
D711	D757	M002	0.406	F_2
D23	D06	UH200	0.500	F_2
D23	D06	UH201	0.500	F_2
0663	1720	D63	0.000	BC_1
0663	1720	D64	0.000	BC_1
0663	1720	D66	0.000	BC_1
0663	1720	D67	0.000	BC_1
0663	D16	D48	0.000	BC_1
0663	D17	D61	0.000	BC_1
0663	1784	D38	0.000	BC_1
0663	0623	D46	0.000	BC_1
0663	D17	D62	0.000	BC_2

Table D. Pedigree relationships among progeny lines of chapter 5 developed from the same parental lines.

No	O1	O2	$f_{(O1,O2)}$
Flint			
1	UH003	F014	0.55
2	UH003	F017	0.55
3	F009	F011	0.58
4	F009	UH002	0.58
5	F011	UH002	0.79
6	F014	F017	0.88
7	F004	UH001	0.94
8	F004	F006	0.94
9	UH001	F006	0.97
Dent			
10	P009	P015	0.55
11	D63	D67	0.88
12	D64	D67	0.88
13	D66	D67	0.88
14	UH200	UH201	0.89
15	M006	M009	0.92
16	D63	D66	0.97
17	D64	D66	0.97
18	D63	D64	0.98


A
D_16
D_35
D_38
D_27
D_24
D_46
D_43
D_47
D_49
D_68
D_36
D_42
D_39
D_19
D_33
D_37
D_28
D_29
D_34
D_41
D_81
D_25
D_23
D_65
D_44
D_79
D_21
D_67
D_70
D_73
D_32
D_72
D_17
D_64
D_51
D_82
F_83
D_18
D_77
F_54
F_02
F_03
F_04
F_59
F_07
F_08
F_05
F_10
F_55
F_06
F_09
F_12
F_56
F_74
F_87
F_11
F_61
F_58
F_66
0.00
2.48
4.96
7.44
9.92
Euclidean Distance

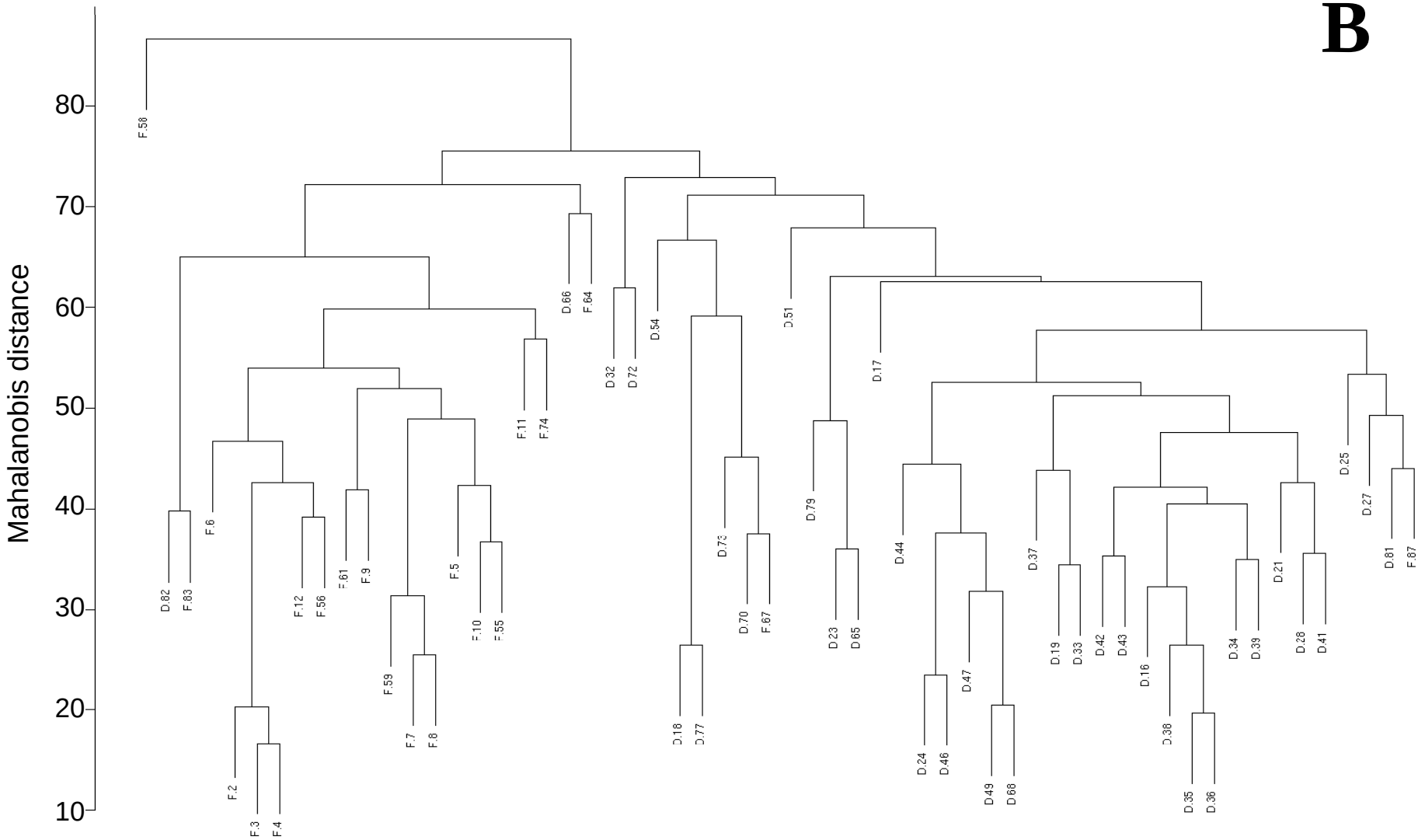


Figure A. Dendrograms based on pairwise (A) Euclidean, or (B) Mahalanobis morphological distances from 25 morphological and agronomic traits. Flint and dent lines are indicated by F and D, respectively.

10. Acknowledgements

First and primarily, I wish to thank my doctoral advisor, Prof. Dr. A.E. Melchinger for his committed guidance and his creative scientific support in the course of the present investigation. Particularly, I am deeply indebted for his trust in me in terms of representing this thesis work on numerous national and international meetings and conferences.

I owe great gratitude to my scientific supervisor, Prof. Dr. Martin Bohn for his committed support within all stages of the work of this thesis. Symbolized by his never closed office door, he had time for my needs whenever I needed his advice. Especially, I highly appreciated the fruitful hints and discussions when I got stuck between endless datasets, countless maize plants or never-ending manuscript corrections.

My thanks are extended to Franz Mauch as well as to Dr. D. Klein and his team of the research station Eckartsweier for the skilled design, realization, and data collection of the field trials. In particular, I owe gratitude to Mr. E. Langmann, Dieter Wiebe, Roland Schoch, Sebastian Pluskat, Maria Wedemeyer, Thomas Schmid, and Adelbert Lehmann for their assistance with a maize field experiment that comprised more than 20% of all maize plots conducted at Eckartsweier over two years. Furthermore, I would like to thank the above-mentioned as well as all unmentioned workers of the research station Eckartsweier for the enjoyable working atmosphere and for their willingness to work overtime during work peaks.

My sincere thanks to Mrs. B. Boesig for being of great help in organizational matters and for being the constant within the everyday life of a plant breeding institute. I am also grateful for the helping hand of Mrs. S. Meyer in financial and itinerary issues.

Special thanks to Prof. Dr. H.F. Utz, Dr. Matthias Frisch, Thomas Magg, Renata Mihaljevic, Jasmina Muminovic, Hans Peter Maurer, Jochen Reif and all unmentioned members within our working group and beyond as well as my office mates Christina Dußle, Marcel Quint, Susanne Dreisigacker, and Pingzhi Zhang for creating a pleasant work environment in combination with the necessary fun during (sometimes extended) working hours and for their support in many scientific and non-scientific issues.

Moreover, I am obliged to our project partners for their essential contributions to studies of this thesis, namely Drs. Jeroen Rouppe van der Voort (Keygene N.V.) and Kader Fatmi (Celera) for their assistance with molecular data, Drs. James Reeves and Vince Lea (NIAB, Cambridge) for their skilled project management, Dr. R. Becher (Bundessortenamt) for his competent advice in terms of observation of morphological traits, and Dr. P. Görtz (Südwestsaat GbR) for the supply with field area at Scherzheim.

Last but foremost, I am deeply indebted to my family, namely Walter, Sofie, and Matthias Heckenberger, my girlfriend Katrin, as well as my friends who supported me in manifold ways, reminded me with great insistence of necessary breaks and rest periods, helped me to keep myself "down-to-earth", and gave me the feeling that there may be certain things that are, in terms of "personal sustainability", at least as important as this thesis.

This research was part of the project "Molecular and other markers for establishing essential derivation in crop plants" supported by a grant from the European Union, Grant No. QLK-CT-1999-01499 (MMEDV). Parts of this research were funded by the 'Gesellschaft zur Förderung der privaten Pflanzenzüchtung' (GFP), Germany as part of the project „Entwicklung von Kriterien zur Definition und Feststellung von abgeleiteten Sorten mittels molekularer Marker", Grant No. A 29/98 HS (97 HS 027).

11. Curriculum vitae

Name:	Martin Heckenberger
Date and Place of Birth:	03.06.1973 in Ochsenhausen
Nationality:	German
Marital Status:	Unmarried
Education	*1980 - 1984*, Grundschule Gutenzell-Hürbel *1984 - 1993*, Gymnasium Ochsenhausen Graduation: Allgemeine Hochschulreife
Military Service	*07/93 – 06/94* at Budel (NL) und Ummendorf (Germany)
Studies	*10/94 – 06/00*, Agricultural sciences at the University of Hohenheim Graduation: Diplom-Agraringenieur
Practical Experience	*07/94 – 10/94*, Alfons Schick, Laupheim, Germany *10/96 – 01/97*, Gordon Bros. Galatea, New Zealand *03/97 – 10/97*, Dr. Reinhard Funk, Gutenzell, Germany
Occupational Activities	*05/96 – 10/96*, Scientific assistant (HiWi) at the Institute for Farm Management, University of Hohenheim, Germany: assignment: linear programming *01/98 – 12/98*, Scientific assistant (HiWi) at the Institute for Crop Production and Grassland Management, University of Hohenheim, Germany: assignment: sample preparation, data processing *09/99 – 06/00*, Scientific assistant (HiWi) at the Institute for Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, Germany: assignment: development of databases with MS Access, data processing
Professional Occupation	*07/00 – 07/03*, Doctorate candidate in Plant Breeding and Applied Genetics (Prof. Dr. A.E. Melchinger), University of Hohenheim, Germany *since 08/03*, Research associate at the Institute for Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, Germany