RESEARCH

Study on Essential Derivation in Maize: III. Selection and Evaluation of a Panel of Single Nucleotide Polymorphism Loci for Use in European and North American Germplasm

Yves Rousselle, Elizabeth Jones, Alain Charcosset, Philippe Moreau, Kelly Robbins, Benjamin Stich, Carsten Knaak, Pascal Flament, Zivian Karaman, Jean-Pierre Martinant, Michael Fourneau, Alain Taillardat, Michel Romestant, Claude Tabel, Javier Bertran, Nicolas Ranc, Denis Lespinasse, Philippe Blanchard, Alex Kahler, Jialiang Chen, Jonathan Kahler, Seth Dobrin, Todd Warner, Ron Ferris, and Stephen Smith*

ABSTRACT

Pairwise distance data for maize (*Zea mays* L.) inbred lines generated using sets of single nucleotide polymorphisms (SNPs) selected from a 50k Infinium array were compared with pairwise distances generated using a set of 163 simple sequence repeat (SSR) loci previously identified to help determine essentially derived variety (EDV) status (UPOV, 1991). Final comparisons were made using 26,874 SNPs after discarding SNPs with insufficient data quality or vulnerability to ascertainment bias. Inbred lines developed in the United States or in western Europe that had been previously published to establish SSR-based thresholds provided the means to determine equivalent SNP-based protocols. Use of 3072 SNPs selected to provide even genomic coverage according to genetic and physical maps provided robust, precise, high discrimination among inbred lines with consistent zonal classification with up to 20% missing data. Comparisons of intercepts and slopes for SSR and SNP inbred pairwise distance data translated the 82% SSR green-orange similarity threshold to 91% using SNPs and the 90% SSR orange-red threshold to 95% using SNPs. Information required to conduct analyses using these 3072 SNPS is presented.

Y. Rousselle and A. Charcosset, INRA, Ferme du Moulon, 91190, Gif-sur-Yvette, France; E. Jones and R. Ferris, Syngenta Biotechnology, 629 Davis Dr., Research Triangle Park, NC 27709; P. Moreau and P. Blanchard, Euralis Semences, Laboratoire de Genetique Moleculaire, Domaine de Sandreau, 6, chemin de Pandedautes, Mondonville, 31705, France; K. Robbins, Dow AgroSciences, 9330 Zionsville Rd., Indianapolis, IN 46268; B. Stich, Max Planck Institute for Plant Breeding Research, Carl-von-Linne-Weg 10, 50829, Cologne, Germany; C. Knaak, KWS Saat AG, Grimsehlstr. 31, 37555, Einbeck, Germany; P. Flament, Z. Karaman, J.-P. Martinant, Limagrain Europe, Laboratoire de genotypage, Batiment 1, La Garenne, Route d'Ennezat, 63720 Chappes, France; M. Fourneau and A. Taillardat, Maisadour Semences, Route de Saint Sever BP27, 40001 Mont de Marsan Cedex, France; M. Romestant and C. Tabel, RAGT Semences, rue Emile Singla, BP3357, F-12033, Rodez Cedex 9, France; J. Bertran, N. Ranc and D. Lespinasse, Syngenta Seeds SAS, 12, chemin de l'Hobit, BP 27, 31790, Saint-Sauveur, France; A. Kahler and J. Kahler, Biogenetics Services,47927 213th St. Aurora, SD, 57002; J. Chen, Agreliant Genetics LLC, 1122 East 169th St., Westfield, IN 46074; S. Dobrin, Monsanto Company, 3302 SE Convenience Blvd., Ankeny, IA 50021; T. Warner, Syngenta Seeds Inc., 317 330th St., Stanton, MN 55017; S. Smith, DuPont Pioneer, 7300 NW 62nd Ave, Johnston, IA 50131. Received 17 Sept. 2014. Accepted 24 Dec. 2014. *Corresponding author (Stephen.Smith@Pioneer.com).

Abbreviations: ASTA, American Seed Trade Association; EDV, essentially derived variety; ISF, International Seed Federation; NAM, nested association mapping; MAF, minimum allele frequency; MRD, modified Roger's Genetic Distance; PIC, polymorphic information content; PVP, plant variety protection; SEPROMA, French Association of Maize Breeders; SNP, single nucleotide polymorphism; SSR, simple sequence repeat; UFS, Union Française des Semenciers; UPOV, Union Internationale pour la Protection des Obtentions Vegetales.

Published in Crop Sci. 55:1170–1180 (2015).
doi: 10.2135/cropsci2014.09.0627
© Crop Science Society of America | 5585 Guilford Rd., Madison, WI 53711 USA

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.

PIONEER 2027
Inari Agric. Inc. v. Pioneer Hi-Bred Int'l, Inc.
PGR2024-00023
1/11

As of July 2014, 70 countries and 2 organizations, the European Union (28 member States) and the African Intellectual Property Organization (17 member States), had become members of the International Union for the Protection of New Varieties of Plants (UPOV) (UPOV, 2014). The UPOV provides a sui generis system to protect intellectual property vested in plant varieties that are declared as distinct, uniform, and stable (DUS) through the grant of Plant Variety Protection (PVP) or Plant Breeders' Rights (PBR) and as implemented in national and/or regional legislation. Of these members, 51 (72%) provide protection under the Act of 19 Mar. 1991, of the Convention. A major distinguishing feature of the 1991 Act compared to previous Acts (Act of 1961 amended in 1972 and Act of 1978) is the addition of the concept of an EDV. An EDV is defined by UPOV (1991) as "predominantly derived from an initial variety, or from a variety that is itself predominantly derived from the initial variety, whilst retaining the expression of the essential characteristics that result from the genotype or combination of genotypes of the initial variety." A major impetus leading to the introduction of the concept of essential derivation was to provide a balance of intellectual property rights (IPR) to breeders of initial varieties and to subsequent breeders or biotechnologists who develop a variety that was improved agronomically, albeit by making relatively small genetic changes to the initial variety. The concept of essential derivation should also support IP by reducing incentives to make minor genotypic or phenotypic changes to an initial variety that result in cosmetic changes with no significant agronomic improvements because breeders of initial varieties would have no incentives to provide commercial licenses for plagiarism (Sanderson, 2006; ISF, 2012). While UPOV provides guidelines for the implementation of the EDV concept (UPOV, 2009) and provides a forum for reviewing relevant research via meetings of the UPOV Biochemical and Molecular Techniques (BMT) working group (UPOV BMT 2007, 2010; UPOV BMT Addendum 2007) it has been largely left to experts in the field, that is, breeders and those who characterize plant varieties, to develop crop specific guidelines or for courts to set legal precedence (District Court, 2008; Court of Appeal, 2009; UPOV BMT 2010; 2011). In contrast, Australia requires the PVP Office to determine EDV status and the goals of EDV implementation in that country is to prevent copying or plagiarism rather than to facilitate balanced opportunities for initial and subsequent breeders to share commercial returns (PBR, 1994).

It is generally accepted that molecular marker-based comparisons of the initial variety and a putative EDV can provide pertinent information to help determine EDV status through an ability to measure genetic conformity (Heckenberger et al., 2002; 2003; Rodrigues et al., 2008; Smith et al., 2014). Sanderson (2006) describes the potential legal and judicial complexities in determining EDV status noting that quantitative data providing measures of genetic similarities between varieties can play an important role in the process. The International Seed Federation (ISF) has promulgated crop specific guidelines to help determine EDV status where an important role of marker data is to determine whether the burden of proof should be reversed to then rest on the developer of the putative EDV (ISF, 2004a; 2004b; 2005, 2006, 2007a; 2007b; 2008, 2009; 2014, nd; UPOV BMT, 2010). A previous study reported the selection and evaluation of a panel of 285 SSR loci to help determine essential derivation in maize (*Zea mays* L.) in the United States (Kahler et al., 2010). A similar study conducted by the French Maize Breeders Association (SEPROMA) resulted in the publication of a set of 163 SSRs that were recommended for use to help determine EDV status in maize germplasm in France (Andreau et al., 2003; Heckenberger et al., 2003; Kahler et al., 2010). Thresholds of SSR marker-based similarity were jointly proposed by the cultivar variety subcommittee of ASTA and by SEPROMA to categorize putative EDVs in relation to an initial variety. For there to be the possibility of essential derivation, the putative EDV had to be more than 82% similar to the initial variety. Similarities above 82% were sufficient to reverse the burden of proof to the breeder of the putative EDV. Similarities above 90% could be regarded as a strong indication of EDV status (ISF, 2008; 2014). Since that time, SNPs have largely replaced SSRs as the marker system of choice for use in plant breeding due to their availability in great numbers, ability to provide higher map resolution, higher-throughput, lower cost, lower error rate, and greater potential to provide harmonized data across laboratories due to their greater clarity in scoring and databasing (Hamblin et al., 2007; Jones et al., 2007; Inghelandt et al., 2010; Yang et al., 2011). It is therefore important that technical guidelines to help determine EDV status with SNPs in maize are available.

The SNPs and SSRs have different attributes which result in contrasting capabilities to estimate genetic distances. These must be taken into account when determining protocols that will provide for equivalent results in helping to determine EDV status. The SSRs generally reveal a higher mean number of alleles. The SNPs generally have a maximum of two alleles per locus and can therefore provide a maximum gene diversity measure or polymorphic index content (PIC) of 0.5. In contrast, SSRs can approach a PIC of 1.0 due to their multi-allelic nature and because they have generally been subject to more cycles of selection for loci that are polymorphic (Hamblin et al., 2007). For example, Hamblin et al. (2007) found that 89 SSRs performed better at distinguishing among maize inbred lines (mostly unrelated by pedigree) than did 847 SNPs. Yu et al. (2009) determined that more than 10-fold the number of SNPs would be required to provide an equivalent ability to distinguish among maize

genotypes. Yang et al. (2011) determined that 884 SNPs provided an equivalent ability to distinguish among maize inbred lines as did 82 SSRs. Inghelandt et al. (2010) surveyed 1537 elite maize inbred lines using 359 SSRs and 8244 SNPs and found a range of 2 to 53 (mean 14.57) alleles per SSR compared to two alleles per SNP. Most studies reported above used <1000 SNPs and showed measures of SSR- and SNP-based genetic distances were only well correlated for subsets of closely related inbred lines. In contrast, when Inghelandt et al. (2010) surveyed 1537 elite maize inbred lines using 359 SSRs and an order of magnitude more (8244) SNPs, then correlations for pair-wise genetic distances according to marker class were evident for both related and unrelated inbred lines. In making such comparisons it is important to recall that selection of a set of markers to be highly discriminative in one set of germplasm due to their high PIC values can lead to ascertainment bias rendering that set of markers potentially less effective or biased at discriminating among other inbred lines that are unrelated to the inbred set used to select the high PIC markers (Hamblin et al., 2007).

By virtue of their mean lower number of alleles, SNPs generally reveal increased allele frequencies and in comparison to multi-allelic marker systems, a decreased proportion of polymorphic markers that are common between two individuals. Consequently, SNP-based comparisons lead to decreased estimates of genetic distances compared to those calculated using SSRs. An imperfect correlation between genetic distances calculated by each marker class also can be expected because different marker systems can sample the genome unevenly. Also, the mutation rate of SSRs is much higher (1 × 10–9) than for SNPs (1 × 10–5) (Yang et al., 2011). Consequently, SSR-based comparisons of germplasm either increase apparent genetic distances due to mutation to a new state or they can reduce genetic distances by mutation leading to homoplasy (Rafalski and Tingey, 2008). The level of precision or coefficient of variation (CV) is also an important consideration. Inghelandt et al. (2010) determined that for 1537 elite maize inbred lines encompassing four unrelated heterotic groups that 90 SSR or 650 SNPs (1:7 ratio) were required to provide a stabilized plateau for CV. Under the assumption that a CV of 1% is sufficient, Inghelandt et al. (2010) estimated that 270 SSRs or 3150 SNPs (1:11 ratio) would suffice, in concordance with theory (Laval et al. (2002) and Vignal et al., 2002) and empirical research (Yu et al., 2009). With further regard to sampling strategy, Inghelandt et al. (2010) and Nelson et al. (2011) showed that slightly lower CVs could be obtained by choosing markers that were equally distributed through the genome on comparison to their random selection according to map position. Stratified sampling therefore allows for increased efficiencies in either the need to select fewer markers or to maximize precision (Inghelandt et al., 2010; Nelson et al., 2011).

Marker-based distances are more appropriate than pedigree records to assess genetic relationships among maize inbreds (Inghelandt et al., 2010; Smith et al., 2014). Both SSR and SNP data can provide the same portrayal of the structure and diversity of maize germplasm (Inghelandt et al., 2010). Nonetheless, due to their differential characteristics, results from empirical studies must be examined before SSR-based guidelines, including marker-based thresholds, can be translated into SNP-based protocols that provide equivalent outcomes in helping to determine EDV status (Inghelandt et al., 2010). We also wished to identify a set of SNPs that could provide a robust, unbiased, and universal method to estimate genetic distance across a broad spectrum of maize germplasm. To help achieve this objective, we started with a large number of SNPs. We then proceeded to exclude from further analyses those SNPs that had already been highly selected according to their ability to reveal polymorphisms between a narrow base of inbred lines (B73 and Mo17) to avoid potential ascertainment bias. Subsequent selections were made with the primary purpose of maintaining even genomic and physical map coverage without selection for PIC to thereby minimize potential limitations imposed by ascertainment bias. Three components required consideration. First, to determine the number of SNPs required to provide acceptably high levels of correlation and low CV. Second, to measure the slope for SSR-based distances compared to SNP-based distances to determine the conversion formula to provide their equivalent threshold levels as previously determined using SSR data. Third, to determine which genetic map (genetic, physical, or both) to use as the basis for marker selection. It was jointly agreed among the maize research membership of the ASTA and the French Breeders Association (UFS) to implement research to determine SNP-based thresholds equivalent to those already established using SSRs (Andreau et al., 2003; ISF, 2008; Kahler et al., 2010). The objectives of this research are to (i) report on the comparison of pairwise distance data among inbred lines of maize generated using SSR and SNP data and (ii) the use of those data to derive SNP-based methods that can help determine EDV status in maize.

MATERIALS AND METHODS

Materials

Maize genotypes used to compare SSR and SNP data are presented in Tables 1 and 2. Inbreds listed in Table 1 had previously been included among those to determine SSR-based EDV thresholds (Andreau et al., 2003; ISF, 2008; Kahler et al., 2010). These inbreds (Table 1) were used to compare correlations of pair-wise distances for SSR and SNPs, to determine numbers of SNP loci that would provide adequate precision, and to provide an initial estimate of SNP thresholds that would provide equivalent results in helping to determine EDV status

Table 1. List and description of inbreds included in the present study to provide the basis for recalibration for essentially derived variety (EDV) thresholds to single nucleotide polymorphism (SNP) data.

Inbred lines developed in the United States			
Inbred	Family	Pedigree	Breeder
PH207	Iodent	G3BD2 × G3RZ1	Pioneer Hi-Bred
A632	B14	(Mt42 × B14) s B14 (3)	University of Minnesota
B14	B14	BSSS	Iowa State University
B37	B37	BSSS	Iowa State University
PHB47	B37	B37 (2) × SD105	Pioneer Hi-Bred
B73	B73	BSSS	Iowa State University
CM105	B14	V3 × B14 (2)	Morden, Canada
CO109	CO109	Early Butler	Ottawa, Canada
PHG29	Iodent	PH207 × (PH207 × 806)	Pioneer Hi-Bred
PHG35	Misc.	G3BD2 × PH595	Pioneer Hi-Bred
PHG39	B14-B37	A33GB4 x A34CB4	Pioneer Hi-Bred
PHG39	Oh43	PH041 × MKSDTE C10	Pioneer Hi-Bred
PHG84	Misc.	PH848 × PH595	Pioneer Hi-Bred
DKHBA1	Misc.	Pioneer brand hybrid 3195 × Pioneer brand hybrid 3199	Pioneer Hi-Bred
I205	Iodent	Iodent	Iowa State University
PHJ40	Misc.	PHB09 × PHB36	Pioneer Hi-Bred
LH1	B37	(B37 × 644) × B37	Holden Foundation Seeds
LH123HT	Misc.	Pioneer brand hybrid 3535	Holden Foundation Seeds
LH132	B73	(H93 × B73) × B73	Holden Foundation Seeds
LH146HT	B14-B73	(B73 × CM105) × CM105	Holden Foundation Seeds
LH74	B14-B73	A632 × B73	Holden Foundation Seeds
LH82	Iodent	610 × LH7	Holden Foundation Seeds
LH82	Misc.	Pioneer brand hybrid 3978	Holden Foundation Seeds
Mo17	C103	C103 × C.I. 187-2	University of Missouri
Mo17	Oh43	W8 × Oh40B	Ohio State University
Mo17	OTHER 2	C.I.540 × Ill. L	Ohio State University
DKPB80	B73	(1067-1 × B73) × B73	DeKalb Genetics
W117	OTHER 3	643 × Minnesota 13	University of Wisconsin
W153R	OTHER 3	(Ia153 × W8) × Ia153	University of Wisconsin
PHZ51	Misc.	PH814 × PH848	Pioneer Hi-Bred

Inbred lines developed in France	
Inbred	Pedigree background
L138	EU Flint
L67	EU Flint
L93	EU Flint
L20 = F2	EU Flint
L48	Lancaster Sure Crop
L89 = W401	Lancaster Sure Crop
L134	Lancaster Sure Crop
L14 = LH59	Lancaster Sure Crop
L102	Group 3
L90	Iodent
L88	Iodent
L130	Iodent
L96 = MBS847	Iodent
L71	Iodent
L50	Iodent
L2	B14-B73
L133	B14-B73
L85	B14-B73
L37 = LH132	B73
L28	B14-B73
L13 = B73	B73
L115	B14-B73
L27	B14-B73
L76 = A632	B14
L1	B14-B73

as SSR data (Kahler et al., 2010). An additional set of proprietary closely related pairs (Table 2) of inbred lines developed either in the United States or western Europe, with pedigrees known only by the respective company representatives, were also then examined using the selected SNP loci to validate and, if necessary to further adjust SNP thresholds to provide greatest consistency with SSR-based categorization of inbred pairs according to possible EDV status.

Simple Sequence Repeat and Single Nucleotide Polymorphism Profiling Protocols and Quality Control Criteria for Single Nucleotide Polymorphism Data

All inbreds (Tabled 1and 2) were profiled using (i) 150 ASTA designated SSRs according to methods reported by Kahler et al. (2010) and (ii) 163 SSRs designated by SEPROMA according to SSR EDV guidelines (Andreau et al., 2003; ISF, 2008). There were 40 SSRs in common between the profiling conducted by Biogenetics and by Trait Genetics. The acrylamide gel technologies used by Trait Genetics (ISF, 2008) were more discriminative among SSR alleles compared to the agarose technology used by Biogenetics (Kahler et al., 2010). Consequently, data from each SSR set were treated as separate datasets.

All inbreds (Tabled 1 and 2) were profiled using a 50K SNP Illumina Infinium array by Trait Genetics GmbH (Gatersleben, Germany) using methods described by Ganal et al. (2011). This set of SNPs was selected to have a minimum allele frequency (MAF) of >0.2 among 25 diverse maize inbreds that were used for Nested Association Mapping (NAM) and which were selected on the basis of their agronomic importance in the United States or to capture as much of the genetic diversity present in maize as possible (Gore et al., 2009; Ganal et al., 2011). The SNP data were first subject to a quality assurance

Table 2. List and background information for additional closely related inbred lines used in the present study to further validate the single nucleotide polymorphism (SNP)-based essentially derived variety (EDV) thresholds. The percent by pedigree for additional pedigree data are minimum estimates of pedigree relatedness because the percent pedigree relatedness among parent lines has not been taken into account.

Breeder	Inbred	Adaptation maturity	Genetic family	Additional pedigree data
DuPont Pioneer	PHB1			
DuPont Pioneer	PHB2			
DuPont Pioneer	PHK4.2			
DuPont Pioneer	PHT5.5			
Syngenta	SYT1.0			
Syngenta	SYT1.1			
Syngenta	SYT2.0			
Syngenta	SYT2.1			
Syngenta	SYT2.2			
Dow Agrosciences	DAS3.0			
Dow Agrosciences	DAS3.1			
Dow Agrosciences	DASS3.1			
Dow Agrosciences	DAS3.2			
Dow Agrosciences	DAS4.0			
Dow Agrosciences	DAS4.1			
Monsanto	MON8.1			
Monsanto	MON8.2			
Monsanto	MON8.3			
Monsanto	MON8.4			
Limagrain	UFS1 EU	Flint zone B–C	Flint	
Limagrain	UFS2 EU	Flint zone B–C	Flint	
Limagrain	UFS3 EU and US	Dent 95–100 d	Iodent-BSSS	
Limagrain	UFS4 EU and US	Dent 95–100 d	Iodent-BSSS	
KWS	UFS5 EU	Zone B–C	Flint	
KWS	UFS6 EU	Zone B–C	Flint	50% by pedigree UFS5
KWS	UFS7 EU	Zone A–B–C	Dent	
KWS	UFS8 EU	Zone A–B–C	Dent	75% by pedigree UFS7
RAGT	UFS9 EU	Dent 85–95 d	B14-B73	
RAGT	UFS10 EU	Dent 85–95 d	B14-B73	50% by pedigree UFS9
RAGT	UFS11 EU	Dent 85–95 d	Iodent	
RAGT	UFS12 EU	Dent 85–95 d	Iodent	UFS11 and UFS12 are sibs
Euralis	UFS13 EU and US	Zone A–B–C	Flint-Iodent	
Euralis	UFS14 EU and US	Zone A–B–C	Flint-Iodent	
Euralis	UFS15 EU and US	Zone A–B–C	BSSS	
Euralis	UFS16 EU and US	Zone A–B–C	BSSS	75% by pedigree UFS15
Maisadour	UFS17 EU	Flint zone A–B	European flint	
Maisadour	UFS18 EU	Flint zone A–B	European flint	
Maisadour	UFS19 EU	Dent 95–105 RM	Lancaster Sure Crop	
Maisadour	UFS20 EU	Dent 95–105 RM	Lancaster Sure Crop	75% by pedigree UFS19
Caussade	UFS21	Dent 95–105RM	Iodent-BSSS	
Caussade	UFS22	Dent 95–105RM	Iodent-BSSS	
Caussade	UFS23	Flint Zone B–C	Flint	
Caussade	UFS24	Flint Zone B–C	Flint	75% by pedigree UFS23

analysis. The SNP data were eliminated from subsequent analyses if they exceeded the following thresholds:

>10% of missing SNPs per genotype
>10% of missing genotype per SNP
>5% heterozygous SNPs per genotype
>10% heterozygous genotypes per SNP

Data from a total of 5882 SNPs were discarded because they did not meet quality standards. A further 1948 SNPs were monomorphic and so data obtained using those SNPs were discarded from subsequent analyses. A total of 14,870 SNPs were from a set of SNPs that had been selected purely on the basis that they were polymorphic between B73 and Mo17. These SNPs were shown to distort genetic distance estimates at least for Mo17 among a broad range of maize germplasm (Ganal et al., 2011) due to ascertainment bias and were therefore discarded from further analyses. Final data used to evaluate important parameters associated with determination of EDV status including to select SNP sets were obtained from a set of 26,874 SNPs.

Measurement and Comparison of Genetic Distances

Genetic distances among pairs of inbred lines were computed using Roger's distance. This method was used instead of the modified Roger's distance method adjusted by Wright (1978) that we reported in Kahler et al. (2010) because it is preferred (Rogers, 1991) and to align with more recent analyses performed in Europe under the auspices of SEPROMA. The modified Roger's distance is simply the square of unadjusted Roger's distances (Reif et al., 2005).

Comparisons of genetic distances between pairs of inbred lines using SSR or SNP data were made by measuring their correlations using Mantel tests and coefficient of determination of least squared regression. Initial comparisons of SNP data were made using subsets comprised of 384, 768, 1536, and 3072 SNPs. For each SNP number class, 100 independent subsets of SNPs were sampled for analysis. Standard deviations and coefficients of variation for pair-wise genetic distances for each pair of inbreds were calculated for each of 100 random subsets at each SNP number class and for the complete data set of 26,874 SNPs.

Determining the Number of Single Nucleotide Polymorphisms

The number of SNPs that would be required to generate pairwise inbred genetic distances with high repeatability and high precision was determined by comparing correlations and standard deviations of distance measures for different sized subsets of SNPs (384, 768, 1536, 3072) compared to the full set of SNPs (27,874). Subsets of markers were chosen only to maintain even coverage across the genome using either physical or genetic maps. There was no selection for PIC to avoid potential future pitfalls associated with ascertainment bias.

Selection of Single Nucleotide Polymorphisms to Comprise the Essentially Derived Variety Set in Relation to the Physical and Genetic Map

We present below a narration of the thought process and strategies that we used to determine a method of selecting specific SNPs according to the physical and/or genetic maps. Random selection of 1536 SNPs from among the 26,874 set resulted in an average marker distance of 1338,238 bp with SD 1514,978 bp compared to an average marker distance of 1362,192 with SD of 610,351 when 1536 SNPs were selected. Selection of SNPs on map information allowed for more even genomic distribution and thus a more even sampling of the maize genome. The question then arose as to how SNPs should be selected; using the genetic map, the physical map, or a combination of both maps. Application of the EDV concept can help prevent plagiarism of an existing genotype. In this respect then genetic distance (recombinations per generation) would be the preferred selection criterion. On the other hand, compared to the genetic map the distribution frequency of expressed sequence tags (ESTs) on the physical map would be expected to be higher in the centromeric and pericentromeric regions due to relatively lower recombination events in these chromosomal regions. It may be therefore that centromeric and pericentromeric regions might be more important with regard to phenotypic expression of genes than previously considered (Gent et al., 2012). Consequently, if the selection of EDV SNPs were to be based on genetic distance alone, then there is a possibility that a subsequent breeder might be able to use SNPs located in the telomeric regions to select genetically away from the initial inbred, at least according to comparisons provided by SNPs, while retaining sufficient genotype to copy the phenotypic expression developed by the initial breeder. It is well established that changes in gene spacing, relative location, and gene presence occur in maize (Anderson et al., 2006). Consequently, it was decided to use both the physical and genetic maps and to select SNPs in equal numbers using each map.

A genetic algorithm was used to select SNP panels based on both genetic and physical distances. The genetic algorithm is a commonly used method to search for optimal solutions (Mitchell, 1996), in this case sets of SNPs that find the optimal balance of genetic and physical distances between selected markers. The algorithm was initiated with SNP panels selected by using bins based on the genetic and physical maps, respectively. The bins had a fixed width, calculated to yield 1536 evenly spaced bins containing two or more SNP makers with VeraCode quality scores >0.9. Two SNP markers were then randomly selected from each bin to yield a 3072 SNP marker panel for genetic and physical distances. The two panels were treated as "chromosomes" and recombination events were simulated to generate new panels. The new panels were then evaluated and selected according to the following fitness function:

$$\text{fitness} = \text{std}_{pd} + \text{std}_{gd} + \text{abs}(\text{std}_{pd} - \text{std}_{gd})$$

where std_{pd} is the standard deviation of the scaled physical distance intervals between selected markers and std_{gd} is the standard deviation of the scaled genetic distance intervals between selected markers. The distance intervals were scaled by dividing them by the total map length. The selected panels or "chromosomes" were then crossed and new recombination events were generated. This process was repeated until the panel became fixed and no further improvement could be made.

Robustness-missing Data Analysis

To examine the impact of missing data on calculations of genetic distance and EDV classification, code was developed to randomly select a predetermined percentage of markers to treat as missing. After removing markers calls for the selected SNPs, genetic distances were recalculated for all pairwise combination of lines and compared to the full set of 3072 markers. This process was repeated 100 times for each selected level of missing data.

Recalibration of Simple Sequence Repeat-based Essentially Derived Variety Zonal Thresholds to Single Nucleotide Polymorphism-Based Essentially Derived Variety Thresholds, Final Threshold Selection, and Validation

The SNP-based EDV thresholds were recalibrated from SSR-based thresholds (ISF 2008; Kahler et al., 2010) by multiplying

the SSR threshold genetic distances of 0.18 (82% similar) and 0.10 (90% similar) for the green–orange and orange–red thresholds, respectively by the slope generated by comparing the UFS 163 SSR pairwise distances with those generated using SNP pairwise distances for the same set of inbred lines. An initial calibration factor was calculated using the 26,874 SNPs. A final calibration factor was calculated once the number and identity of a specific SNP EDV set had been determined and tested for robustness. The SNP-based thresholds were then validated by comparisons of zonal EDV classifications for additional pairs of proprietary inbred lines (Table 2) using the UFS 163 SSR data and SSR-based thresholds with those generated using the recommended set of 3072 EDV SNPs and SNP-based EDV thresholds.

RESULTS

Characteristics of the Various Datasets

Genetic diversity (*He*) averaged across the 150 ASTA designated SSRs using agarose gels (Kahler et al., 2010) calculated using the 90 inbreds (Tables 1 and 2) was 0.55 and 0.67 when averaged across the 163 UFS SSRs using acrylamide gels. Pair-wise inbred genetic distances ranged from 0.05 to 0.70 using the ASTA 150 inbreds and agarose gel technology (Kahler et al., 2010) and from 0.02 to 0.88 using the UFS SSRs and gel technology. The UFS methodology revealed greater SSR allelic diversity amongst the set of inbreds. Consequently, further analyses comparing SSR data were with pair-wise distances generated using the UFS methodology.

Correlations of Genetic Distances between Pairs of Inbred Lines as Determined using Simple Sequence Repeats Compared to 26,874 Single Nucleotide Polymorphisms

The correlation between pairwise genetic distances estimated with UFS SSRs using acrylamide gels and the 26,874 SNPs was highly significant ($P_{value} < 2.2.\ 10^{-16}$), $R^2 = 0.92$; slope of the regression was 0.50 (Fig. 1).

Determining Parameters for the Number of Single Nucleotide Polymorphisms for Use in Essentially Derived Variety Determination and Comparisons with Simple Sequence Repeat Data

We selected a mean genetic distance of 0.1 to be able to compare error rates for each SNP set, as this distance was anticipated to be close to the threshold for EDV determination. Standard deviations for a mean genetic distance of 0.1 for 384, 468, 1536, and 3072 subsets of SNPs were 0.012, 0.008, 0.006, and 0.004, respectively. Coefficients of variation for a mean genetic distance of 0.1 for these same SNP subsets were 0.15, 0.10, 0.07, and 0.05, respectively. In contrast, and under the simplifying assumption that each SSR marker is independent, standard deviation for the genetic distance of 0.2 (equivalent to 0.1 as estimated by SNPs) was 0.031 with a coefficient of variation

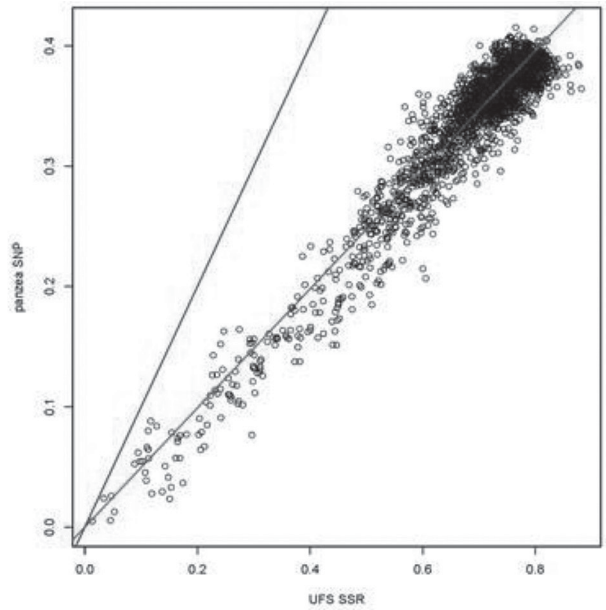

Figure 1. Correlation between genetic distances among pairs of 90 inbred lines (Table 1) estimated with Union Française des Semenciers (UFS) simple sequence repeat (SSR) and single nucleotide polymorphism (SNP). Blue line is the first bisector and red line the regression line.

of 0.16 using the UFS 163 SSR set. According to this comparison as few as 384 SNPs provided a more accurate estimation of genetic distance than did 163 SSRs. Correlations between pairwise distances estimated using the different subsets and the complete set of 26,874 SNPs were $R^2 = 0.95$ (384 subset), $R^2 = 0.98$ (768 and 1536 subsets) and $R^2 = 0.99$ (3072 subset).

Determining the Distribution of Single Nucleotide Polymorphisms

Given the large number of SNP markers available, we initially hypothesized that random selections of marker sets would provide the basis for adequately even genomic coverage. We tested this hypothesis by comparing randomly selected SNPs with those selected on the basis of providing for even physical map distribution. We found that SNP sets selected to provide even genomic coverage gave more uniform genome distribution as determined by smaller SD of average marker intervals (SD 610,351 bases for selected sets compared with SD 1514,978 bases for random selections). As described in the Materials and Methods, we embarked on selecting SNP sets that optimized even map coverage based on both physical and genetic map distances. The developed algorithm successfully minimized standard deviations of the marker intervals on the genetic and physical maps resulting in an average distance standard deviation of 1.2 cM for marker intervals on the genetic map and 1182,338 bases on the physical map (compared with 2 cM on the genetic map if selection was based on physical map,

and 3240,229 bases on the physical map if selection was based on the genetic map).

Final Selection of Single Nucleotide Polymorphisms and Recalibration of Single Nucleotide Polymorphism-Based Thresholds

The initial basis of recalibrating the SSR-based EDV thresholds to their equivalence based on SNPs was founded on a comparison of 26,874 SNPs with UFS 163 SSR pairwise distance data that showed a correlation R^2 of 0.92 with a slope of 0.5 (Fig. 1). Based on analyses of correlations with SSR data and error rates associated with estimated genetic distances, 3072 SNPs were selected to assist in EDV determination. The 3072 SNPs are identified in Supplemental Table S1 together with their genetic and physical map locations. The 3072 SNP set consisted of two subsets of 1536 SNPs both selected based on their genetic and physical genome distribution. Pair-wise genetic distances calculated using either subset of 1536 SNPs and the combined set of 3072 were highly correlated (R^2 = 0.98. 0.99, and 0.99, respectively) with those calculated using the set of 26,874 SNPs. Dividing the 3072 SNP set into two equal subsets gives the opportunity for users to be able to gather initial estimates of genetic distance using one smaller 1536 SNP set that can then be followed up, if for example greater precision is required, using the complete set of 3072 SNP markers. Using the combined set of 3072 SNPs the correlation of all pairwise distances for 90 inbred lines with pairwise distances computed using the UFS standard set of SSRs was R^2 = 0.94 with a slope of 0.5 (Fig. 2). This was the highest R^2 value found for all SNP sets that we examined using various selection criteria (either physical or genetic map data or at random with regard to map location) when compared to pairwise distances computed using the USF SSR set.

Recalibration of Simple Sequence Repeat-based Essentially Derived Variety Thresholds to Single Nucleotide Polymorphism-based Thresholds

The slopes resulting from comparing pairwise genetic distances for the same inbred lines (Tables 1 and 2) using UFS 163 SSRs and either 26,874 SNPs or the specifically recommended set of 3072 SNPs were 0.5. Consequently, the SSR maize green–orange EDV threshold of 0.18 genetic distance (82% genetic similarity) translated to 0.18 × 0.5 = 0.91 genetic distance (91% genetic similarity).. Likewise, the SSR orange–red threshold of 0.10 genetic distance (90% similarity) was recalibrated to 0.05 genetic distance (95% similarity). European and U.S. lines were also analyzed separately to determine whether these thresholds were appropriate within each germplasm pool; green–orange and orange–red thresholds were 91.1 and 95% for U.S. lines and 90.7 and 94.8 for European lines, respectively.

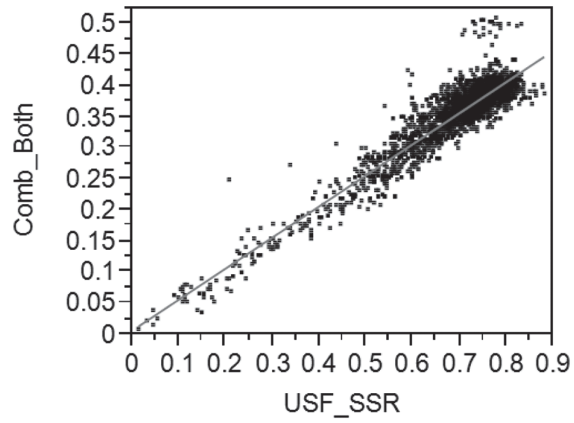

Figure 2. Correlation of pairwise distances between 90 inbred lines of maize included in the present study (Table 1a) on the basis of 3072 single nucleotide polymorphisms (SNPs) selected following quality control by analyzing jointly physical and genetic map data compared with pairwise distances computed using 163 Union Française des Semenciers (UFS) selected simple sequence repeats (SSRs), R^2 = 0.93 with a slope of 0.5.

Usage of 91% (green–orange) and 95% (orange–red) SNP thresholds resulted in fewer re-categorizations when either 1536 or 3072 SNPs were used (10–15/2198 or 0.5% inbred pairs) compared to application of 90% (green–orange) and 94% (orange–red) thresholds (15–17/2198 or 0.7–0.8% inbred pairs). Very importantly, transitioning the methodology and classification of inbred pairs from previously accepted SSR threshold to these SNP thresholds caused no inbred pairs to be reclassified across two non-adjacent zones (i.e., red to green or vice versa).

Robustness in the Face of Missing Data

As percent missing data rose from 1 to 50% then correlations of pairwise distance data compared to that obtained using the full set of data fell from 0.999 to 0.994 and SDs rose from <0.0001 to 0.0085. Of greatest relevance were resultant changes of zonal categorizations of lines (Fig. 3). Categorization of inbred lines was robust for adjacent changes with <0.5% inbreds changing into adjacent zones with up to 20% missing data. The level of missing data had to reach 45% before there was any change in classification of an inbred line across two zones (green–red).

DISCUSSION

The EDV concept is an important component of the protection of intellectual property provided by Plant Variety Protection or Plant Breeders' Rights as legislated in accordance with guidelines promulgated by the UPOV 1991 Convention (UPOV, 1991). Practical application of the EDV concept encourages the breeding of new distinct varieties that have significantly improved agronomic characteristics but in other respects may be closely similar to the initial variety. When a variety is determined to be essentially derived from an initial variety, it is the sole

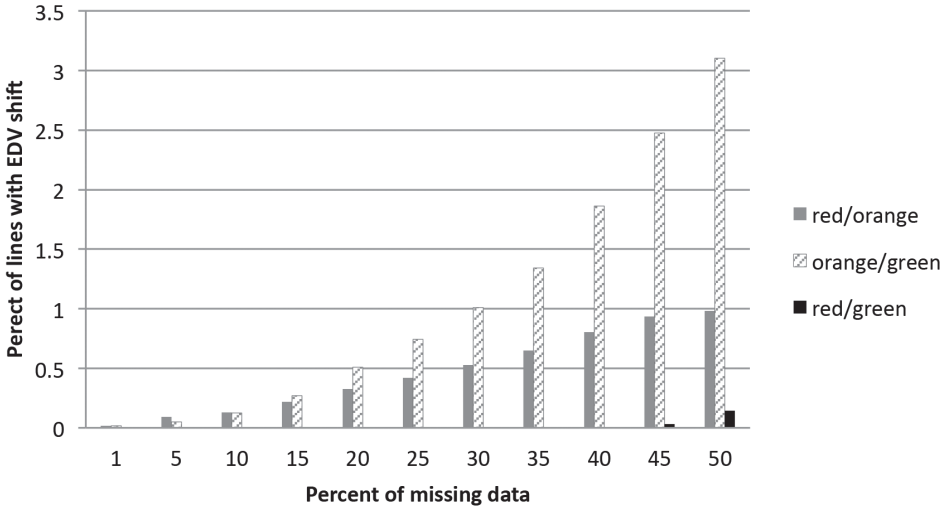

Figure 3. Percentages of inbred lines with shifts in zonal classification with subsets of single nucleotide polymorphisms (SNPs) containing up to 50% missing data compared to using 3072 SNPs.

responsibility of the breeder of the initial variety to decide whether to allow the second breeder to commercialize the resulting EDV. When commercialization of an EDV is agreed then this will usually be under terms where both breeders can be rewarded for their respective contributions. The EDV concept also provides protection against plagiarism, cosmetic breeding, or counterfeiting that results in germplasm that is closely similar to an existing variety or parental inbred line protected with a PBR but which otherwise has no significant additional agronomic merit (Sanderson, 2006; ISF, 2012).

However, it is often difficult for the breeder of the initial variety to assemble the information required to determine EDV status. It is particularly difficult to make morphological comparisons of parental lines of hybrid crops because parental lines are usually maintained as trade secrets. The ISF (UPOV BMT, 2010; ISF, 2012) (http://www.worldseed.org/isf/edv.html) has therefore proposed that genotypic data should be used to help provide an initial evaluation of the likelihood that a subsequently developed inbred line, variety, or hybrid might be an EDV. If an inbred or variety is shown to be a putative EDV then ISF (2012) proposes that there should be a reversal of the burden of proof from the claimant (initial breeder) to the second breeder. Sanderson (2006) describes the potential complexities in determining EDV status noting that while quantitative data may not alone be able to provide the final basis for determination of EDV status, those data play an important role in the process.

There are numerous publications supporting the use of molecular data to provide measures of genetic similarity for inbred lines and varieties and their usefulness to help determine EDV status (Smith et al., 2014). The ISF has published methodologies for several crops (ISF, 2004a; 2004b; 2005, 2006, 2007a; 2007b; 2008, 2009; 2014) including for maize using SNPs. Several authors have compared the abilities of SSRs and SNPs to provide comparisons of maize germplasm that are meaningful in terms of pedigrees, their genetic constitution, and assignment into various heterotic or familial groups including Van Inghelandt et al. (2010) and Nelson et al. (2011). These authors concur that, so long as adjustments are made that take into account the contrasting characteristics of the different marker systems, either marker method can be usefully applied to characterize and distinguish maize germplasm. Methods have already been developed for the use of SSRs to help determine EDV status in maize (Andreau et al., 2003; ISF, 2008; Kahler et al., 2010). The technical advantages associated with SNPs compared to SSRs are leading many laboratories to transition to the use of SNPs (Vignal et al., 2002, Jones et al., 2007, Yang et al., 2011). In addition Inghelandt et al. (2010) indicated "that under the assumption of a fixed budget, modified Roger's distance (MRD) and *gene diversity* (*D*) could be more precisely estimated with SNPs than with SSRs."

Inghelandt et al. (2010) noted that use of SNP data leads to decreased genetic distance estimates compared to those estimated using SSRs and that consequently "new thresholds have to be defined, if essentially derived varieties will be, in the future, identified based on SNP markers instead of SSR or restriction fragment length polymorphism markers (cf. ISF, 2008)". We therefore undertook this study with the initial purpose of evaluating the use of SNP markers in comparison to SSR markers in estimating genetic distances between pairs of inbred lines of maize that were selected to represent U.S. and EU germplasm including lines that had previously been used to estimate SSR-based thresholds for the determination of EDVs (Andreau et al., 2003; ISF, 2008; Kahler et al., 2010). We found high correlation of genetic distances between SNP-based measures and SSR-based measures, especially when more discriminative polyacrylamide gel-based methods were used to distinguish among SSR alleles

(Andreau et al., 2003). We examined subsets of SNPs and established that 1536 SNPs gave a level of correlation and precision (R^2 = 0.98 SD 0.006) compared to the 24,886 set of SNPs that would be anticipated to be sufficient for the purposes of helping to determine EDV status. Nonetheless, since costs associated with the use of SNPs can be expected to further decrease we decided to identify two sets of 1536 SNPs which when combined provided a correlation of R^2 = 0.99; SD of 0.004 compared to use of the 26,874 SNPs. We found that selection of SNPs based on genetic and physical distances gave best distribution on both genomes and gave the highest correlations with SSR data. Genetic distances estimated using SNP data were generally lower than those estimated using SSRs. Consequently, equivalency to a predetermined SSR threshold requires that the SNP threshold to be raised in relation to the slope generated by comparing pairwise distances from inbred lines using SSR and SNP-based methods. As a result of this recalibration we concluded that an SSR-based maize EDV threshold of 82% similarity (ISF, 2008) should be recalibrated to ≥91% similarity using SNPs (ISF, 2014). Likewise, an SSR threshold of 90% similarity (ISF, 2008) should be recalibrated to ≥95% when using SNPs (ISF, 2014).

We tested robustness of zonal classification of inbreds using these SNP protocols in the face of missing data. Classifications were robust with up to 25% missing data (Fig. 3), well above a 1.5 to 5% level of missing data that might be expected (Jones et al., 2007; Simic et al., 2009; Yan et al., 2010). It is important to note that use of 50% missing data is not equivalent to use of one of the two 1536 SNP sets, for each 1536 SNP set was selected to allow even sampling of the genome according to both the genetic and the physical maps, therefore giving lower error rates and SD compared with random missing data. We decided not to select subsets of SNPs based on their PIC value among U.S. and European inbred lines. We concluded that use of 3072 SNPs selected to provide even coverage of the genetic and physical maps from among the larger set of 26,874 SNPs provided adequate power and precision to measure genetic distances with a sufficient CV to help determine EDV status. In addition, the cost efficiencies that now allow 3072 SNPs to be assayed on a regular basis obviated the potentially negative aspects associated with increasing ascertainment bias that could result from a more stringent and targeted selection of SNPs. Being able to avoid problems associated with ascertainment bias therefore makes this a potentially useful marker set for use with other germplasm unrelated to that tested here. We also draw attention to the derivation of the set of SNPs included in this study which spanned a very broad base of *Zea* germplasm. More than half are tropical in origin, nine are temperate lines, two are sweet corn lines (representing Northern Flint), and one is a popcorn inbred line (McMullen et al., 2009). We anticipate therefore that the set of 3072 SNPs published here may be more broadly useful in characterizing and distinguishing among maize germplasm that is distributed more globally. This set of SNPs can also provide the pool from which further subsets of SNPs could be chosen for use, at least in the United States and Europe, for the purposes of maize inbred line or hybrid identification, pedigree verification, and to help in the management of reference collections which is an integral part of the distinctness, uniformity, and stability (DUS) examination process required to determine eligibility for the receipt of protection by PVP. We therefore recommend that other national/regional breeders also evaluate this set or subsets of SNPs for use in the identification of maize inbred lines or hybrids, varietal purity, and to help in the determination of EDV status.

Supplemental Information Available

Supplemental information is available with the online version of this manuscript.

Acknowledgments

We wish to acknowledge the administrative assistance of Candace Madsen and data analyses provided by Delphine Madur and Stephane Nicolas.

References

Anderson, L.K., A. Lai, S. Stack, C. Rizzon, and B. Gaut. 2006. Uneven distribution of expressed sequence tag loci on maize pachytene chromosomes. Genome Res. 16:115–122. doi:10.1101/gr.4249906

Andreau, B., P. Dubreuil, D. Perret, F. Ananza, and A. Charcosset. 2003. Synthesis of studies conducted by SEPROMA on the estimate of genetic distance between maize inbred lines. INRA, Station de Genetique Vegetale Ferme de Moulon, F-91190, Gif sur Yvette, France, and SEPROMA, 17 rue du Louvre, F75001, Paris. Available upon request from UFS, 17 rue du Louvre, F75001, Paris.

Court of Appeal. 2009. Judgment in the case of Danziger "Dan" Flower Farm versus Astee Flowers BV. Court of Appeal, The Hague, the Netherlands.

District Court. 2008. Van Zanten Plant BV v. Hofland BV Summary. The Hague, the Netherlands

Ganal, M.W., G. Duretwitz, A. Polley, A. Berard, E.S. Buckler, A. Charcosset et al. 2011. A large maize (*Zea mays* L.) SNP genotyping array: Development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS ONE 6:1–15 e28334. doi:10.1371/journal.pone.0028334

Gent, J.I., Y. Dong, J. Jiang, and R.K. Dawe. 2012. Strong epigenetic similarity between maize centromeric and pericentromeric regions at the level of small RNAs, DNA methylation and H3 chromatin modifications. Nucleic Acids Res. 40:1550–1560. doi:10.1093/nar/gkr862

Gore, M.A., J.-M. Chia, R.J. Elshire, Q. Sun, E.S. Ersoz, B.L. Hurwitz et al. 2009. A first-generation haplotype map of maize. Science 326:1115–1117. doi:10.1126/science.1177837

Hamblin, M.T., M.L. Warburton, and E.S. Buckler. 2007. Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. PLoS ONE 12:e1367. doi:10.1371/journal.pone.0001367

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. doi:10.1023/A:1020539330957

Heckenberger, M., J.R. 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: Genetic and technical sources of variation in AFLP data and comparison with SSR data. Mol. Breed. 12:97–106. doi:10.1023/A:1026040007166

Inghelandt, D.V., A.E. Melchinger, C. Lebreton, and B. Stich. 2010. Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theor. Appl. Genet. 120:1289–1299. doi:10.1007/s00122-009-1256-2

ISF. 2004a. Guidelines for the handling of a dispute on essential derivation in lettuce. Int Seed Federa, Nyon, Switzerland.

ISF. 2004b. Technical protocol for implementation of the ISF guidelines for the handling of a dispute on EDV in lettuce. Int Seed Federa, Nyon, Switzerland.

ISF. 2005. Essential derivation information and Guidance to Breeders. Int Seed Federa, Nyon, Switzerland.

ISF. 2006. Use of DNA Markers for DUS Testing, Essential Derivation and Identification. Int Seed Federa, Nyon, Switzerland.

ISF. 2007a. Guidelines for the Handling of a Dispute on Essential Derivation in Cotton. Int Seed Federa, Nyon, Switzerland.

ISF. 2007b. Guidelines for the Handling of a Dispute on Essential Derivation in Oilseed Rape. Int Seed Federa, Nyon, Switzerland.

ISF. 2008. Guidelines for the Handling of a Dispute on Essential Derivation of Maize Lines. Int Seed Federa, Nyon, Switzerland.

ISF. 2009. Guidelines for Handling a Dispute on Essential Derivation in Ryegrass. Int Seed Federa, Nyon, Switzerland.

ISF. 2012. ISF View on Intellectual Property. Int Seed Federa, Nyon, Switzerland. http://www.worldseed.org/cms/medias/file/PositionPapers/OnIntellectualProperty/View_on_Intellectual_Property_2012.pdf.

ISF. 2014. ISF Guidelines for handling disputes on essential derivation of maize lines. Int. Seed Federa, Nyon, Switzerland.

ISF. nd. Explanatory Notes: Regulation for the Arbitration of Disputes concerning Essential Derivation (RED) Int Seed Federa, Nyon, Switzerland. http://www.worldseed.org/cms/medias/file/Rules/EssentialDerivation/Explanatory_Notes.pdf

Jones, E.S., H. Sullivan, D. Bhattramakki, and J.S.C. Smith. 2007. A comparison of simple sequence repeat and single nucleotide polymorphism marker technologies for the genotypic analysis of maize (*Zea mays* L.). Theor. Appl. Genet. 115:361–371. doi:10.1007/s00122-007-0570-9

Kahler, A.L., J.L. Kahler, S.A. Thompson, R.S. Ferriss, E.S. Jones, B.K. Nelson et al. 2010. North American study on essential derivation in maize: II. Selection and evaluation of a panel of simple sequence repeat loci. Crop Sci. 50:486–503 10.2135/cropsci2009.03.0121. doi:10.2135/cropsci2009.03.0121

Laval, G., M. San Cristobal, and C. Chevalet. 2002. Measuring genetic distances between breeds: Use of some distances in various short term evolution models. Genet. Sel. Evol. 34:481–507. doi:10.1186/1297-9686-34-4-481

McMullen, M.D., S. Kresovich, H.S. Villa, P. Bradbury, H. Li, Q. Sun et al. 2009. Genetic properties of the maize nested association mapping population. Science (Washington, DC) 325:737–740.

Mitchell, M. 1996. An introduction to genetic algorithms. MIT Press, Cambridge, MA.

Nelson, B.K., A.L. Kahler, J.L. Kahler, M.A. Mikel, S.A. Thompson, R.S. Ferriss et al. 2011. Evaluation of the numbers of single nucleotide polymorphisms required to measure genetic distance in maize (*Zea mays* L.). Crop Sci. 51:1470–1480. doi:10.2135/cropsci2010.07.0401

PBR. 1994. Plant breeder's rights act 1994. IP Australia, Gov. of Australia, Canberra.

Rafalski, A., and S. Tingey. 2008. SNPs and their use in maize. In: R.J. Henry, editor, Plant genotyping II- SNP technology. CABI, Cambridge, MA. p. 30–43.

Reif, J.C., A.E. Melchinger, and M. Frisch. 2005. Genetic and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management. Crop Sci. 45:1–7. doi:10.2135/cropsci2005.0001

Rodrigues, D.H., N.F. de Alcantara, and I. Schuster. I 2008. Identification of essentially derived soybean cultivars using microsatellite markers. Crop Breed. Appl. Biotechnol. 8:74–78.

Rogers, J.S. 1991. A comparison of the suitability of the Rogers, modified Rogers, Manhattan, and Cavalli-sforza and Edwards distances for inferring phylogenetic trees from allele frequencies. Syst. Biol. 40:63–71. doi:10.1093/sysbio/40.1.63

Sanderson, J. 2006. Essential derivation, law and the limits of science. Law in Context, Vol. 24. Research Paper Series 09-16, Soci-legal research centre, Griffith Law School, Griffith Univ., Brisbane, Queensland, Australia. p. 34–53.

Simic, D., T. Ledencan, A. Jambrovic, Z. Zdunic, J. Brkic, A. Brkic et al. 2009. Genetika 41:237–246. doi:10.2298/GENSR0903237S

Smith, J.S.C., E.S. Jones, and B.K. Nelson. 2014. The use of molecular marker data to assist in the determination of essentially derived varieties. In: R. Tuberosa, et al., editors, Genomics of plant genetic resources. Springer, Dordrecht, the Netherlands. p. 50–65.

UPOV. 1991. International convention for the protection of new varieties of plants. UPOV, Geneva, Switzerland. http://www.upov.int/en/publications/conventions/1991/act1991.htm (accessed 19 Feb. 2015).

UPOV. 2009. Explanatory notes on essentially derived varieties under the 1991 act of the UPOV Convention. UPOV/EXN/EDV/1. UPOV, Geneva.

UPOV. 2014. International union for the protection of new varieties of plants: Publ. 437. (In English.) Geneva http://www.upov.int/export/sites/upov/about/en/pdf/pub437.pdf (accessed 19 Feb. 2015).

UPOV BMT. 2007. EDV in corn: Concepts of essential derivation and dependence; Possible use of DNA markers: The maize case. Ad hoc crop subgroup on molecular techniques for maize, Chicago, 3 December. BMT-TWA/Maize/2/7-a. UPOV, Geneva.

UPOV BMT Addendum. 2007. Guidelines for the handling of a dispute on essential derivation of maize. Addendum to document BMY-TWA/Maize/2/7/-a. BMT-TWA/Maize/2/7/-a Add, Chicago, 3 December. UPOV, Geneva.

UPOV BMT. 2010. EDV–The ISF approach. BMT/12/22. Twelfth session. Ottawa, Canada. 11–13 May 2010. UPOV, Geneva.

UPOV BMT. 2011. An EDV court case in wheat in Germany. Presentation by Marcel Bruins, Secretary General ISF at the UPOV BMT meeting, November 2011. UPOV BMT, Brasilia, Brazil.

Vignal, A., D. Milan, M. SanCristobal, and A. Eggen. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genet. Sel. Evol. 34:275–305. doi:10.1186/1297-9686-34-3-275

Wright, S. 1978. Evolution and the genetics of populations, Vol. 4. Variability within and among natural populations. Univ. of Chicago Press, Chicago.

Yan, J., X. Yang, T. Shah, H. Sanchez-Villeda, J. Li, M. Warburton, Y. Zhou, J. Crouch, and Y. Xu. 2010. High-throughput SNP genotyping with the GoldenGate assay in maize. Mol. Breed. 25:441–451. doi:10.1007/s11032-009-9343-2

Yang, X., Y. Xu, T. Shah, H. Li, Z. Han, J. Li, and J. Yan. 2011. Comparison of SSRs and SNPs in assessment of genetic relatedness in maize. Genetica (The Hague) 139:1045–1054. doi: 10.1007/s10709-011-9606-9.

Yu, J.M., Z.W. Zhang, C.S. Zhu, D.A. Tabanao, G. Pressoir, et al. 2009. Simulation appraisal of the adequacy of number of background markers for relationship estimation in association mapping. Plant Gen. 2:63–77. doi:10.3835/plantgenome2008.09.0009

View publication stats