DOI: 10.1002/csc2.20201

ORIGINAL RESEARCH ARTICLE

Crop Breeding & Genetics

Single nucleotide polymorphisms facilitate distinctness-uniformity-stability testing of soybean cultivars for plant variety protection

F. Achard¹ | M. Butruille¹ | S. Madjarac¹ | P.T. Nelson¹ | J. Duesing² | J-L. Laffont³ | B. Nelson³ | J. Xiong³ | Mark A. Mikel⁴ ^(b) | J.S.C. Smith⁵ ^(b)

¹ Bayer Crop Science, 700 Chesterfield Pkwy W, Chesterfield, MO 63017, USA

² West Des Moines, IA 50265, USA

³ Corteva Agriscience, 7000 NW 62nd Ave, Johnston, IA 50131, USA

⁴ Department of Crop Sciences, University of Illinois, Carl R. Woese Institute for Genomic Biology, 1206 West Gregory Avenue, Urbana, IL 61801, USA

⁵ Department of Agronomy, Iowa State University, 716 Farm House Ln, Ames, IA 50011, USA

Correspondence

J.S.C. Smith, Department of Agronomy, Iowa State University, 716 Farm House Ln, Ames, IA 50011. Email: stephen.smith@mchsi.com

Assigned to Associate Editor Aaron Lorenz.

Bayer Crop Science purchased Monsanto Corporation June 2018. Corteva Agriscience™ was formed June 2019 following the 2017 merger of Dow and DuPont-Pioneer.

Abstract

Plant variety protection (PVP), or plant breeders' rights, provides intellectual property protection (IPP) for cultivars. Technical requirements are distinctness, uniformity, and stable (DUS) reproduction. However, field trials are increasingly resource demanding and potentially inconclusive for soybean (Glycine max [L.] Merr.). Our objective was to establish methodologies using molecular markers to facilitate DUS testing while maintaining current IPP levels. We determined that DNA from 10-15 bulked plants represented cultivar genotype. Single nucleotide polymorphism (SNP) data were highly robust in the face of missing and mistyped data; concordances among five laboratories were >.9888. We used SNP, morphological, physiological, and pedigree information to examine 322 publicly available cultivars including 187 with PVPs. Associations among cultivars following multivariate analyses of genetic distances from SNP data and from pedigree kinship data were very similar. A SNP similarity of 98.6% was the maximum at which cultivars also differed for morphological characteristics. Many (38%) cultivar pairs with members >90% SNP similarity expressed different morphologies with SNP similarities ranging 96–98.6%. Of cultivars <96% SNP similar, only a single pair differed by a single morphological difference; all others differed by more than two morphological characteristics. A SNP similarity of 96% between soybean cultivars represents an initial and conservative point of demarcation between cultivars that have morphological differences and those that do not. Chronological monitoring of pedigree-kinship and SNP similarities showed little evidence that

Abbreviations: CP, coefficient of parentage; DUS, distinctness, uniformity, stable; GRIN, Germplasm Resources Information Network; IPP, intellectual property protection; MAF, minor allele frequency; PVP, plant variety protection; SNP, single nucleotide polymorphism; SP, single plant; UPOV, International Union for the Protection of New Varieties of Plants.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. Crop Science published by Wiley Periodicals LLC on behalf of Crop Science Society of America

a lack of genetic diversity in F_2 breeding populations contributed to challenges in DUS among U.S. soybean cultivars.

1 | INTRODUCTION

The development and release of a new soybean cultivar takes 6-10 years (Fehr, 1978; Jamali, Cockram, & Hickey, 2019; Scaboo, Chen, Sleper, & Clark, 2017). If breeders wish to recoup investments, they obtain IPP on new varieties (Blair, 1999; Lence, Hayes, Alston, & Smith, 2015). Intellectual property protection is important to obtain for varieties developed by commercially funded breeders (Blair, 1999; Thomson, 2013) and has increasingly become usual practice for publicly funded breeding programs in the United States (Shelton & Tracy, 2017). The most common approach to obtain IPP is through PVP, also known as plant breeders' rights. Most countries have adopted the sui generis system established by the International Union for the Protection of New Varieties of Plants (UPOV) (http:// www.upov.int). Plant variety protection provides exclusive time-limited ownership rights for the sale and repeated use of cultivars and parental lines of hybrids.

Technical requirements for a cultivar to be granted PVP are distinctness from all other publicly known varieties of the crop species and a level of uniformity consistent with the biology of reproduction and maintenance strategy required to allow stable reproduction of cultivars within that species. These are collectively known as the DUS requirements. This DUS testing involves comparisons of morphologically expressed characteristics. The UPOV (1998a; 1998b) lists 20 morphological characteristics for DUS testing of soybean; however, individual PVP authorities can request additional information. The Community Plant Variety Office of the European Union requests data for 18 morphological characteristics (Community Plant Variety Office, 2017), while the U.S. PVP Office specifies 19 morphological characteristics. The U.S. PVP Office also requests "any available information on reaction states" for causal organisms of three bacterial diseases, 12 fungal infections, five viral diseases, seven nematodes, three insects, seven herbicides, and further information on nearly 100 pathogenic races, six physiological reactions, seven herbicides and six seed composition characteristics (https://www.ams.usda.gov/sites/default/files/media/02-Soybean%20ST-470-02%202015.pdf). The Argentinean Instituto Nacional de Semillas requests information for 36 morphological characteristics and reactions to two bacteria, 18 fungi, three viruses, three nematodes, two insects, six specified herbicides plus others, and three specified seed composition characteristics and others.

The Brazilian Serviço Nacional de Proteção de Cultivares requests information for 38 characteristics.

Many factors limit the power of morphological characteristics in their usage to determine distinctness (Jamali et al., 2019). Genotype \times environment (G \times E) interactions markedly affect expression of morphological characteristics (Khan, Khalil, & Taj, 2003; Liu et al., 2017a; Staub, Gabert, & Wehner, 1996; Wurtenberger, 2017) thereby reducing precision and, consequently, their discriminatory power. Not all character states are found in equal frequency thereby further reducing discrimination power (Kumar, Rani, Jha, Rawal, & Husain, 2017; Law et al., 2011). For example, most U.S. soybean cultivars express broad (ovate) leaflets as opposed to narrow or lanceolate leaflets (Dinkins, Keim, Farno, & Edwards, 2002). Power of distinction is yet further reduced by correlations in expression states of different characteristics thereby reducing the number of different combinatorial character states (Law et al., 2011). For example, expression for days to flowering and days to maturity, plant height, branches per plant, pods per plant, seeds per pod, and seed weight are correlated (Malek, Rafii, Afroz, Nath, & Mondal, 2014) as are hypocotyl color and flower color (Ramteke & Murlidharan, 2012). Morphological expression fails to reveal underlying genotypic differences. For example, genetic networks that contribute to hilum color are involved in expression of flower color, pubescence color (Palmer, Pfeiffer, Buss, & Kilen, 2004), and stem termination (Bandillo et al., 2017), resulting in fewer observable and discriminating expressed character states than the diversity of their underlying genetic mechanisms (Bandillo et al., 2017; Fang et al., 2017).

There are additional challenges that impinge upon the use of morphological characteristics in DUS testing. Field-based DUS trials are very labor intensive and expensive (da Silva et al., 2017; Hariprasanna, 2018; Rathinavel, Manickam, & Sabesh, 2005; Staub et al., 1996; Tommasini et al., 2003; UPOV, 2015; UPOV TWC, 2010; Wagner & McDonald, 1981) and ultimately depend upon an element of subjectivity (Jarman & Hampson, 1991; Staub et al., 1996; Tommasini et al., 2003; Singh et al., 2004; Heckenberger, Bohn, Klein, & Melchinger, 2005; Karivaradaraaju, 2005; Kumar, 2014; Hariprasanna, 2018; Gopal et al., 2018). Soybean reference collections are large (Song et al., 1999) and expand annually (Jones, Jarman, Austin, White, & Cooke, 2003). As of 2013, 831 soybean varieties had been registered in Argentina at an average annual rate of 44 during 2010-2013 (Craviotti, 2015). During 2000-2009 the U.S. PVP Office received an average of 52 applications per year for new soybean varieties. However, during 2010-2018 the annual number of new soybean applications had more than tripled to 162 (https://apps.ams.usda.gov). There are currently >4000 soybean cultivars in the U.S. reference collection including publicly developed cultivars that were not tested for PVP and 2617 other cultivars with PVP issued from July 1975 to November 2018 (USDA, 2019). As of 2013, there were >1000 soybean varieties registered in Brazil with ~600 cultivars protected by the National Cultivar Protection Service (Ribeiro, Tanure, Maciel, & de Barros, 2013). Numbers of soybean varieties for which applications for protection were sought rose from seven in 1997 to an average of 55 per year during 1998-2011 (Santos, de Moraes Aviani, Hidalgo, Machado, & Araújo, 2012), further increasing to nearly 100 per year during 2014-2017 (Campante, 2018), indicating that the rate of increase can reach several hundred per annum (McDonald, 1984; Oda et al., 2015; UPOV, 2005). As of 2017 there were 2030 varieties of soybean cultivated in China (Liu et al., 2017a).

Consequently, as the numbers of candidate and publicly known cultivars increases, the ability to distinguish among them all on the basis of morphological traits alone becomes more difficult even though differences in agronomic performance may exist (Hariprasanna, 2018; Lombard, Baril, Dubreuil, Blouet, & Zhang, 2000; McDonald, 1984). For example, difficulties in establishing distinctness on the basis of morphological characteristics have been reported from Argentina (Giancola, Lacaze, & Hopp, 2002), Brazil (Boldt, Sediyama, Nogueira, Matsuo, & Teixeira, 2007; Dos Santos Silva et al., 2016; Nogueira et al., 2008; Vieira, Pinho, Carvalho, & Silva, 2009), India (Kumar et al., 2017), and the United States (Adams, 1996: Diwan & Cregan, 1997; Rongwen, Akkaya, Bhagwat, Lavi, & Cregan, 1995). In these circumstances, initial morphological comparisons are increasingly liable to fail to provide a sufficient basis to evaluate distinctness unless they are later augmented with additional morphological and physiological data, which then requires more time and resources to complete the examination process (Hariprasanna, 2018; Lombard et al., 2000; McDonald, 1984). With increased usage of new breeding technologies, it is anticipated that the ability to distinguish among cultivars using current DUS criteria will become even more difficult (UPOV, 2016a). In contrast, Diwan and Cregan (1997) and Yoon et al. (2007), were rapidly able to discriminate among 36 U.S. soybean cultivars that "were seemingly identical based upon maturity, seed coat color, hilum color, cotyledon color, leaflet shape, flower color, pod color, pubescence color and plant habit" (Yoon et al., 2007) on the basis of comparing the SSR and SNP profiles of those cultivars, respectively.

3

Both applicants and PVP agencies incur significant costs in the design, planting, monitoring, and analyses of data from replicated field trials in attempts to address G×E interactions (Comstock & Moll, 1963; Camussi, Spagnoletti Zeuli, & Melchiorre, 1983; Patterson & Weatherup, 1984; Staub et al., 1996; Lombard et al., 2000; UPOV, 2002, 2019; Singh et al., 2004; Law et al., 2011; Ojo, Ajaya, & Oduwaye, 2012; Ramteke & Murlidharan, 2012; Korir et al., 2013; Kumar, 2014; Oda et al., 2015; Kumar et al., 2017; Ranatunga, Arachchi, Gunasekare, & Yakandawala, 2017; Wurtenberger, 2017; Gopal et al., 2018). For example, two cycles of field trials and data analyses are required for most species with management and data collection costs per cycle reported in the Netherlands of EUR€1855-2530 (US\$2041-2783), totaling EUR€3710-5060 (US\$4081-5566) per cultivar (USDA-Agricultural Marketing Service, 2016) (exchange rate 25 Oct. 2019). Additional time and resources are required if field trials are of insufficient quality because of weather (e.g. drought, storms, flooding) or other unforeseen circumstances. Meanwhile, as resources required for testing increase, implementing agencies are under pressure to become more cost-effective (Pourabed et al., 2015; Bundessortenamt, 2017).

The use of molecular marker data can contribute to DUS testing by virtue of their: (a) high discriminatory power, (b) high repeatability, (c) freedom from $G \times E$ interaction (Noli, Teriaca, Sanguineti, & Conti, 2008), (d) applicability to seed or early growth stages of plants (Jamali et al., 2019), (e) speed of data production and analysis, (f) continued reduction in costs, and (g) amenability to readily searchable databases comprising records for thousands of cultivars, which can also (h) facilitate global harmonization (De Riek, 2001; van Ettekoven, 2017).

The UPOV (2011; 2016b) has given a positive assessment for the use of molecular marker data in DUS testing: Model 1, "Molecular characteristics as a predictor of traditional characteristics or use of molecular characteristics which are directly linked to the traditional characteristics (gene specific markers)" and Model 2, "Calibration of threshold levels for molecular characteristics against the minimum distance in traditional characteristics." Model 1 can be difficult to apply (Cockram, Jones, Norris, & O'Sullivan, 2012), as a result of the biological challenges and resources required to identify marker-trait associations, which can maintain their robustness across diverse germplasm. Model 2 potentially provides the basis for the introduction of a "system for combining phenotypic and molecular distances in the management of variety collections" as a means to improve speed and efficiency of distinctness evaluation (Norris, Jones, Cockram, Smith, & Mackay, 2012). However, the suitability of an approach founded on this model can be elusive because it is dependent upon a high correlation of morphological characteristics and molecular marker data in their abilities to differentiate between cultivars (Jones et al., 2013).

Use of either Model 1 or Model 2 in DUS testing requires a crop-species-specific approach, as emphasized by the UPOV Technical Committee, that "use is acceptable within the terms of the UPOV Convention and would not undermine the effectiveness of protection offered under the UPOV system" and that "it is a matter for the relevant authority to consider if the(se) assumptions are met" (UPOV, 2011; 2016b). Single nucleotide polymorphism data are routinely used in the management of reference collections of maize (Zea mays L.) by French PVP authorities, according to Model 2, whereby inbred lines are declared "super-distinct" when SNP-based similarities and singleyear morphological similarities both fall below a certain threshold, thereby eliminating the need for a second season of morphological comparisons (Maton et al., 2014; Thomasset et al., 2015; UPOV, 2011; 2016b).

There are important considerations when using molecular techniques in DUS: (a) to maintain existing levels of IPP (De Riek, 2001), more simply stated as "how different is different?" (Wallace, 2017), stemming from circumstance where molecular data provide greater discrimination than phenotypic comparisons (Terzić, Zorić, & Seiler, 2020); (b) to provide a level playing field for all breeders regardless of their resource capabilities; (c) to make the process more efficient and potentially more harmonized globally; (d) to maintain or reduce cost; and (e) to avoid levels of uniformity that are unrealistic, overly expensive, unnecessary, or impractical to achieve (International Seed Federation, 2012).

We have adopted a phased approach to evaluating the usefulness of molecular markers for distinctness evaluation in soybean, one which takes into account these considerations. Phase 1 involves selection of SNP marker sets and evaluation of variety sampling methodologies for DNA extraction. Phase 2 is focused on the establishment of a SNP-based threshold of pairwise intercultivar similarity, below which soybean varieties can be considered distinct. To accomplish this, we investigated the following. First, we measured levels of SNP intracultivar heterogeneity in the context of established soybean breeding methodologies whereby bulking occurs at or around the F₄ stage. Second, we ascertain the discriminatory capability of SNPs, including comparison to morphological characteristics and pedigree-based coancestry or kinship. Third, we examine the robustness of SNP data with respect to (a) marker number, (b) missing data, (c) scoring error, and (d) interlab repeatability. Finally, we monitored genetic diversity of biparental crosses made to develop segregating populations across a three decadal period to ascertain trends in genetic diversity between parents of breeding crosses and therefore potentially within the resultant F₂ segreACHARD ET AL.

gating breeding populations. The rationale for monitoring diversity being that, if a trend toward diminishing genetic diversity were to be observed, then a reduction in diversity might also be expected to contribute to challenges in establishing distinctness regardless of the type or nature of characteristics being examined.

2 | MATERIALS AND METHODS

2.1 | Germplasm selection

We identified 322 publicly available soybean cultivars from among cultivars developed by either public or proprietary (commercially oriented) breeding programs (Supplemental Table S1). These cultivars collectively represent those that had been important in U.S. soybean production and in further breeding during the 1970s, 1980s, and 1990s. A subset of 187 off-PVP cultivars bred by commercial organizations was also used in select analyses as described below. Cultivars comprising this subset have been granted PVP by the USDA-Agricultural Marketing Service; each has therefore satisfied all DUS requirements. Morphological and pedigree-kinship data are more readily available for the 187 than for the full set of 322 cultivars. Seed for the 322 cultivars is available through the USDA Germplasm Resources Information Network (GRIN) (https://www.ars-grin.gov/) system, including for those comprising the 187 subset per the USDA policy to release off-PVP cultivars into the public domain.

2.2 | Single nucleotide polymorphism set selection

Two iSelect Illumina Infinium BeadChip arrays are publicly available for assaying soybean SNPs: the SoySNP50K (Song et al., 2013; SoyBase, 2018) with 50,000 SNPs and the BARCSoySNP6k with SNPs selected from the SoySNP50K chip by the Soybean Genomics and Improvement Laboratory, Beltsville Agricultural Research Center, MD (Illumina, Inc., 2015; Song et al., 2014). The SoySNP50K and BARCSoySNP6k SNP sets have been used in various mapping and genetic characterization studies (Akond et al., 2013; Gibson, 2015; Huang et al., 2018; Liu et al., 2017b; Urrea, Rupe, Chen, & Rothrock, 2017). The SoySNP50K chip was also used to genotype the full 20,087 USDA soybean germplasm collection (soybean GRIN collection) (Song et al., 2015) and those SNP data were generously made available to the soybean research community (https://soybase.org/dlpages/).

We used publicly available SNP data for analyses using the complete set of 322 cultivars and for those cultivars

5

comprising the 187 subset. Each of these cultivars had the same 5,346 SNPs reported from among the BARC-SoySNP6k. For these cultivars, we used only SNP data for these 5,346 SNPs as reported from SoySNP50K and BARC-SoySNP6k in order to provide a balanced set of SNP cultivar data for analysis. For interlab comparison, seed sampling, and intracultivar heterogeneity analysis, new DNA was extracted, profiled, and scored following de novo genotyping using the entire set of SNP assays arrayed on the BARCSoySNP6k and using seed obtained from the USDA via GRIN.

2.3 | De novo DNA extraction and genotyping

DNA was extracted at the Monsanto laboratory in St. Louis, MO, USA. Approximately 10 mg of leaf tissue was collected from single plants, lyophilized, ground to powder, and transferred into 1.4 ml Matrix tubes in a 96-well rack. DNA extraction was performed by lysis with buffer, precipitation with potassium acetate, collection in binding buffer on a filter plate, followed by two ethanol washes, then elution in HPLC-grade water. DNA concentration was quantified by using the Thermo Scientific NanoDrop process (Thermo Fisher Scientific Corp.) (https://tools.thermofisher.com/content/sfs/brochures/TN52607-E-0914 M-Oligonucleotides-Mweb.pdf); DNA samples were normalized to 50 ng μ l⁻¹ using HPLC-grade water.

Experiments were conducted at the Monsanto laboratory (Ankeny, IA), the DuPont Pioneer laboratory (Johnston, IA), the Dow Agroscience laboratory (Indianapolis, IN), the Eurofins BioDiagnostics laboratory (River Falls, WI), the Geneseek laboratory Neogen (Lincoln, NE). Genotyping was performed using the Illumina Infinium BARCSoySNP6k (Illumina, Inc., 2015) according to the Infinium HD Assay Ultra Protocol using all SNPs. The SNP alleles were called manually using GenomeStudio Genotyping Module version 2011.1 (Illumina, Inc., 2016) by all laboratories except Monsanto who used proprietary software. The SNPs were called only when they exhibited from one to three discrete allele clusters of one or two classes of homozygotes and heterozygotes (if present) with high signal intensity.

2.4 | Intracultivar heterogeneity and seed sampling

Two important considerations in developing a DNA sampling strategy are (a) the number of plants to be assayed per cultivar and (b) whether to use individual plants or bulks thereof. To investigate these variables, we estimated intracultivar heterogeneity among five cultivars that were known from previous analyses to be representative of the upper range of residual heterogeneity. Two replicates each of varieties 9551, 9171, and 9221 were analyzed in the DuPont-Pioneer laboratory and one replicate each of A2396 and A2855 were analyzed in the Monsanto laboratory. Seeds were planted in growth chambers; 17 single plants (SPs) of each cultivar were sampled for each replicate and DNA was extracted independently for each sample. Aliquots of the single extracts were then combined to create seven bulk samples, where each bulk was comprised of equal amounts of DNA from the SP extracts as follows: (a) plants one to five, (b) plants one to seven, (c) plants one to nine, (d) plants one to 11, (e) plants one to 13, (f) plants one to 15, and (g) plants one to 17. Therefore, for each cultivar there were 24 samples; 17 SP samples and seven bulk samples. In total, 192 samples (136 from SPs and 56 from bulks) were generated.

2.5 | Seed-lot heterogeneity

Heterogeneity of seed lots was reported as the percentage SNPs reported as heterozygous in SPs and the percentage SNPs reported as heterogeneous in each bulk sample. Comparisons between true level of heterogeneity as measured from SP data were made with heterogeneity levels reported from bulks comprising those single plants.

For each single seed, heterozygosity was calculated as follows:

$$H = \frac{1}{2} \sum_{i=1}^{N} \alpha_i$$

Where *N* is the number of markers and $\alpha_i = \int 1$ if Allele $1 \neq$ Allele 2 for marker *i*

0 otherwise

For groups of k single seeds (k = 5, 7, 9, 11, 13, 15, 17), heterogeneity was calculated as follows:

$$h_{k} = \frac{1}{N} \sum_{i=1}^{N} \beta_{i}, \text{ where}$$

$$\beta_{i} = \begin{cases} 1 & \text{if for marker } i, \text{ Allele } 1 \neq \text{ Allele } 2 \text{ in at least} \\ & \text{one seedor there is at leastone different} \\ & \text{ideotype formarker } i \\ 0 & \text{otherwise} \end{cases}$$

For the bulks, heterogeneity was calculated as follows:

$$h_B = \frac{1}{N} \sum_{i=1}^{n} \gamma_i \text{, where}$$

$$\gamma_i = \begin{cases} 1 & \text{if Allele } 1 \neq \text{Allele } 2 \text{ formarker } i \\ 0 & \text{otherwise} \end{cases}$$

2.6 | Minor allele frequency

Minor allele frequency (MAF) was examined in order to more completely understand the contribution of specific ranges of allele frequencies to seed-lot heterogeneity and to allow an additional means to determine optimum numbers of individual plants to comprise a sample bulk. The MAF can be understood as the probability that a SP sampled from the same population is heterogeneous regarding a given marker. Then, the probability (p_i) that there is at least one heterogeneous seed in a sample of k plants regarding a given marker i is equal to the following:

 $p_i = 1 - (1 - MAF_i)^k$, where MAF_i is the MAF for marker *i*

Now, let X_i be the random variable having value 1 if there is at least one heterogeneous seed in the sample of k plants for marker i, and is 0 otherwise. Then the distribution of $Y = \sum_{i=1}^{N} X_i$ is Poisson binomial with mean $\mu = \sum_{i=1}^{N} p_i$ and variance $\sigma^2 = \sum_{i=1}^{N} p_i(1 - p_i)$. The heterogeneity rate h_k is therefore given by $h_k = \mu/N$ and its coefficient of variation (CV) by CV = $\sqrt{\sigma^2/\mu}$.

Minor allele frequency of multiple SPs was measured as the lowest allele count divided by total allele count. For example, if five SP read AA, AA, AA, TT, and AA at one locus, then MAF is 0.2 (two Ts = lowest allele by a total of 10 allele count). For each cultivar, MAF of all SPs was computed across all markers to assess the true heterogeneity of each bulk.

2.7 | Determination of the number of plants to sample

Using inputs of the total number of markers in the assay, the number of heterogeneous markers, the MAF range, and the number of plants sampled, a CV curve was graphed to review the precision using Monte Carlo simulation; for each number of plants sampled, a random MAF value within the MAF range was generated for each marker 10,000 times and the mean of the CVs was computed and displayed.

2.8 | Intracultivar single nucleotide polymorphism heterogeneity

Levels of intracultivar heterogeneity were measured using 5346 SNPs for each of 40 soybean cultivars (Supplemental Table S1). Of these, 35 cultivars were proven to meet DUS examination criteria for the purpose of obtaining PVP. An additional five cultivars were developed by publicly funded programs, and while they may have been subject to examination to meet a level of uniformity determined to be necessary for stable varietal reproduction, they had not been subject to DUS examination for the purpose of obtaining a PVP. These cultivars were chosen with input from soybean breeders so that they collectively represented a range of maturities and release dates. The SNP data were collected by the Monsanto and DuPont Pioneer laboratories using bulk samples of 15 individuals per cultivar to maximize resource use efficiencies with the understanding that bulk sampling can slightly underestimate the actual level of heterogeneity (see results from the sampling strategy experiment that used both individual and bulk sampling).

2.9 | Cultivar comparisons using single nucleotide polymorphism, pedigree, and morphology

A pairwise simple genetic similarity was calculated among cultivar pairs, where the count of identical SNP alleles was divided by the total number of SNPs considering only SNPs that were nonmissing for both varieties in the pair (Song et al., 2015). The software package KIN (Tinker & Mather, 1993) was used to calculate coefficient of parentage (CP) (Malécot, 1948). Values for pedigree similarity range from 0% (no, or at least no known pedigree relationship) to 100% similarity (identicality on the basis of known pedigree). Coefficient of parentage is the probability that two alleles at a randomly selected locus are identical by descent. Coefficient of parentage data are calculated based upon assumptions that (a) progeny inherit genes equally from both parents, that is, there is no selection; (b) parents are homozygous; (c) parental ancestors with unknown pedigrees are unrelated; (d) parental (founder generation) ancestors with unknown pedigrees are equally unrelated; and (e) BC₅ or greater derived isolines are considered equivalent to the recurrent parent (Martin, Blake, & Hockett, 1991; Mikel, Diers, Nelson, & Smith, 2010; Sneller, 1994; Van Beuningen & Busch, 1997; Wang & Lu, 2006). With regard to assumption (d), founder generations, though not linked by pedigree, likely have a number of genes that are identical by descent inherited from remote

ancestors. In contrast, measures of similarity using molecular marker data are not subject to the restraints imposed by these assumptions. A pedigree-based or degree-ofkinship difference between cultivars (1–CP) was calculated as the basis to show associations on the basis of pedigree records using multivariate analysis.

We used two approaches to compare genetic (SNPbased) and pedigree-based estimates of intercultivar similarities and comparisons of associations among cultivars in the basis of these two contrasting types of data. One approach was to use tanglegram analysis, which is a means to readily compare two multivariate analyses of associations among entities (deVienne, 2019; Sang-Tae & Donoughue, 2008). In this case the entities are associations among soybean cultivars on the basis of SNP data and associations among those same cultivars on the basis of known pedigrees. The dendextend package simplifies the creation of tanglegrams and their presentation in publication-ready format (Galili, 2015). DeVienne (2019) has questioned whether tanglegram analysis can accurately provide a formal measure of a lack of congruence between different associations by measuring the tangle or cross-associations of entities. However, our purpose in providing a tanglegram is primarily to provide a simple visual means of comparing the multivariate associations of cultivars on the basis of SNP and pedigree-kinship data. The second approach was to compare intercultivar similarities according to SNP genetic and pedigree-based kinship data by correlation analyses. For these correlation analyses, we used all pedigree-based pairwise distances of cultivars, and for the 187 subset of cultivars, subsets of those pedigree data thereby allowing for analyses that included cultivar pairs with different numbers of generations or depth of pedigree information. The rationale for subsetting pedigree-based kinship data was two-fold. First, the primary focus of this research is upon cultivars that are more related, rather than less or unrelated by pedigree, for it is generally the former that are more likely to be similar in the expression of their morphological phenotypes. Second, pedigree-based estimates of kinship tend to become more informative as the number of generations or depth of pedigree increases.

For initial information on morphological and physiological differences between cultivars comprising the 322 set we scanned comparisons citing the most similar cultivar from published PVP certificates. We focused on comparisons of morphological and physiological data for cultivars with SNP similarities >90% made publicly available using theBARCSoySNP6k. Cultivars comprising the 187 plant variety protected subset had more complete databased records of their morphological and physiological attributes by virtue of each having been submitted for DUS examination and having been granted PVP. This dataset also included several cultivars that are not themselves plant variety protected, but which merit inclusion as reference cultivars. We therefore used morphological and physiological data provided by the U.S. PVP Office to focus detailed comparisons among cultivars involving these data with measures of genetic similarity using SNPs and estimates of relatedness using pedigree data. The subset of 187 plant variety protected cultivars with SNP similarities >90% for which morphological and physiological data were provided by the U.S. PVP Office comprised 53 cultivars (28% of the 187 that were plant variety protected).

Distance measures involving comparisons of morphological data among cultivars were computed using each of two methods. Euclidean distances were computed for the morphological data by first normalizing the data by subtracting the trait mean for each data point and then dividing by the standard deviation. Euclidean distances among cultivars using standardized variables were then estimated using the dist function in R (Core Team, 2019). Since Euclidean distance data result from a synthesis of data described in multidimensional space and combining information from all characteristics (Sneath & Sokal, 1962), it was also informative to examine differences between cultivars on a simpler basis. Therefore, we also calculated the number and percentage difference of expressed morphological and physiological characteristics, both individually and combined, between each pair of cultivars. These distance measures differ from the approach taken by French PVP authorities (GEVES) who use the GAIA software (Gregoire, 2003; 2007; Maton et al., 2014; Thomasset et al., 2015; UPOV TWC 2010), which incorporates an additional layer of differential weightings among individual characteristics. The GAIA software, as understood in the context of this subject, is developed by GEVES to measure, assign weighting for each characteristic, compute, and compare total phenotypic distances between cultivars (Gregoire, 2003). Differences are then "summarised in a synthetic value which allow(s) quantification of the size of the difference on a scale that the crop expert can manage and use over years" (Gregoire, 2007).

Conceptually, the approach to determining a threshold of distinctness requires consideration of sources of variation, such as $G \times E$ interactions, operator error, equipment error, and intracultivar variation. A distinctness threshold can then be established by requiring a specified number of standard deviations of error between cultivars. This summary reflects the best practice adopted by UPOV using morphological and physiological characteristics. However, as previously documented, such an approach is fraught with large sources of unexplained variation (error). Nonetheless, during the 1960s, when UPOV was conceived and for several decades thereafter, molecular markers were either not available or not sufficiently discriminative, practical, or cost-effective for use in DUS examination. Consequently, UPOV relied solely on expressed morphological characteristics for DUS examination.

Our analysis builds upon established foundations and takes into full consideration concerns previously expressed about the use of marker data by establishing a SNP-based distinctness threshold through examination of cultivars that have already been declared DUS in the PVP system, that is, using cultivars with expired PVP certificates. The threshold approach also provides a practical approach to for determining a level of uniformity on the basis of SNP data that enables stable reproduction of cultivars. Levels of SNP heterogeneity, which previously resulted from the use of widely accepted breeding and seed bulking practices coupled with morphological evidence of uniformity, provide a threshold of percentage SNP heterogeneity that has proven demonstrably acceptable, routinely achievable, and supports stable seed increase of cultivars.

2.10 | Robustness of single-nucleotide-polymorphism-based measures of intercultivar similarity using publicly available data generated using the BARCSoySNP6k set

For each of the 322 cultivars (Supplemental Table S1), data for subsets of the 5346 SNP set were selected using 2673, 1336, 668, 334, and 167 SNPs. Two SNP selection methods were used to select two different arrays of each subset. For the first array, SNPs were randomly selected without attention to their map location or individual discrimination ability. For the second array, SNP subsets were selected so that both expected heterozygosity value of the full set (0.357) and even genomic coverage were maintained. Genomic coverage was maintained by selecting SNP loci at the extremes of each chromosome, then with each subsetting exercise, removing SNPs in closest proximity with increasing distance between SNPs as numbers of SNPs in each subset were reduced. Care was also taken in selection of SNPs within each subset to maintain a mean heterozygosity of 0.357 within each subset. Mean, minimum, and maximum centimorgan (cM) distances in parentheses between selected SNPs for the 5346 SNPs and the nonrandomly selected array of subsets were as follows: 5346 (0.5, 0, 6.09); 2673 (0.81, 0, 6.09); 1336 (2.01, 0.01, 8.51); 668 (4.05, 0.01, 14.05); 334 (8.16, 0.01, 33.81); and 167 (16.63, 0.01, 59.3). Similarity matrices were calculated for each SNP subset using a simple matching routine computed at the allele level using Python Version 2.7 (https://www.python.org/psf/). Similarity matrices were compared using Mantel test correlations computed using NTSYSPC Version 2.21q (Rohlf, 2008).

2.11 | Concordance across laboratories

Thirty-five cultivars that were individually proven to meet DUS examination criteria for the purpose of obtaining PVP were used (Supplemental Table S1). These cultivars were chosen with input from soybean breeders to collectively represent a range of maturities and release dates. The SNP genotyping was performed by each of five laboratories (Dow, Eurofins, Gene Seek, Bayer and Pioneer). Two DNA samples, each from different SPs of each cultivar, making 68 samples in all, were SNP profiled using the BARCSoySNP6k SNP set as described previously. The SNP profiling was conducted blind with respect to cultivar identity.

The SNP data quality control was performed at two levels—marker and cultivar—following procedures reported by Song et al. (2013). At the marker level, SNPs having heterogeneity and missing data percentages >10% were omitted. This quality control step retained data for 5,103 markers out of the original 6,000. Of the total 807 SNPs removed, 57 failed for heterogeneity only, 781 SNPs failed for missing data rate only, and 59 SNPs failed for both criteria. At the cultivar level, samples with heterogeneity and missing data >10% were omitted, resulting in the exclusion of 25 cultivars from further analysis.

Levels of concordances between laboratories for each sample were calculated as follows:

Concordance
$$(i, j) = \frac{1}{2} \sum_{k=1}^{N} \delta_k$$

where N is the number of markers and $\delta_k =$

- 0 if marker *k* alleles from laboratory *i* are different from laboratory *j*
- 1 if only one marker k allele from
 - laboratory *i* is identical to one allele from laboratory *j*
- 2 if marker *k* alleles from laboratory *i* and laboratory *j* are identical

2.12 | Chronological monitoring of genetic diversity

Cultivars selected to determine whether there was evidence of a narrowing genetic base in terms of being parents to make F_2 segregating populations for further cultivar development are identified in Supplemental Table S1. We examined both pedigree kinship data and percentage SNP genetic similarities between the parents of F_2 breeding populations that resulted in new soybean cultivars over three decades (1970–1999). Each of

TABLE 1 Mean heterogeneity percentages for single plants and bulks comprised of those respective single plants

Seeds	9171 Rep 1	9171 Rep 2	9221 Rep 1	9221 Rep 2 ^ª	9551 Rep 1	9551 Rep 2	A2396	A2835
				%				
1–5 single seeds	1.22	1.22	2.19	2.19	0.20	0.20	3.82	4.40
5 seeds bulk	1.10	0.76	1.97	1.95	0.19	0.19	3.63	3.88
1–7 single seeds	1.22	1.22	2.19	2.19	0.20	0.20	3.82	4.40
7 seeds bulk	1.10	0.73	1.95	-	0.19	0.19	3.55	3.94
1–9 single seeds	1.22	1.22	2.19	2.19	0.20	0.20	3.82	4.42
9 seeds bulk	1.10	0.73	2.03	2.03	0.19	0.19	3.59	3.86
1–11 single seeds	1.22	0.90	2.10	2.07	0.20	0.20	3.76	4.04
11 seeds bulk	1.14	1.05	2.05	1.99	0.19	0.19	3.59	4.04
1–13 single seeds	1.22	0.90	2.15	2.17	0.20	0.20	3.76	4.12
13 seeds bulk	1.14	1.14	1.98	1.77	0.19	0.19	3.65	4.10
1–15 single seeds	1.22	0.90	2.15	2.19	0.20	0.20	3.82	4.22
15 seeds bulk	1.14	1.14	2.03	1.99	0.19	0.19	3.65	4.04
1–17 single seeds	1.22	1.22	2.19	2.19	0.20	0.20	3.82	4.32
17 seeds bulk	1.14	1.14	2.07	_	0.19	0.19	3.55	4.04

^aHeterogeneity data are not displayed for because the seven-plant bulk gave results that showed an obvious sampling error and for the 17-plant bulk because of genotyping failure.

TABLE 2 Analysis of variance table for the data presented in Table 1. The factor effects tested are the sample type (single plant [SP] vs. bulk), the number of plants in the sample, the variety and their two-way interactions

Source	df	Sum of squares	Mean square	F-value	$\Pr > F$
SP vs. bulk	1	0.605	0.605	62.79	<.0001
No. of seeds	6	0.024	0.004	0.42	.8643
Variety	4	214.814	53.704	5572.45	<.0001
(SP vs. bulk) \times (No. of seeds)	6	0.164	0.027	2.84	.0164
(SP vs. bulk) \times variety	4	0.214	0.053	5.54	.0007
(No. of seeds) \times variety	24	0.060	0.002	0.26	.9998
Error	64	0.617	0.010	_	-

these cultivars had been granted PVPs and had thus met DUS criteria following comparisons of morphological characteristics.

3 | RESULTS

3.1 | Sampling protocol study

3.1.1 | Intracultivar heterogeneity in single and multiplant aliquots

Five cultivars were used to investigate heterogeneity within SPs and within multiplant bulks synthesized by combining DNA from SP extractions (Table 1 and 2). The SP heterogeneity ranged from 0.2 to 4.42%. For bulks, intracultivar heterogeneity varied according to cultivar, ranging from 0.19 to 4.04%. Replicate bulks of

varieties 9171, 9221, and 9551, with each made using a second sampling of 17 plants per cultivar, were consistent for percentage heterogeneity. Heterogeneity rates of bulks were slightly (but significantly) lower than for SPs, although they became very close when bulks were comprised of 10 or more individual plants (Tables 1 and 2). Heterogeneity data were not displayed for cultivar 9221, replication two, because the seven-plant bulk gave results that showed an obvious sampling error and for the cultivar 9221 17-plant bulk there was a genotyping failure.

In order to better understand factors contributing to missed heterozygote genotype calls using bulk samples and to provide an additional means to estimate optimum bulk sizes, we also examined the ability to detect heterogeneity in the bulks as described above according to MAF (Figure 1). The range of MAFs was wide, from <10 to 50%, with most SNP loci falling in the range 25–50%. In contrast,



FIGURE 1 The ability to detect heterogeneity in bulks according to minor allele frequency (MAF)

the count and range of MAF for cultivar 9551 was low and narrow (40–50%). The MAF profiles were consistent across available replications (cultivars 9171, 9221, and 9551). In Figure 2, CV is plotted against the number of plants sampled and the inflection of this curve falls between eight and 12 plants. With bulks comprising 15 plants or more, CVs are stabilized.

3.2 | Intracultivar heterogeneity in multiplant bulks

Among the 36 cultivars sampled (Supplemental Table S1), the highest levels of intracultivar SNP heterogeneity were found for two cultivars that had not been through the DUS examination process for PVP ('Essex' 6% and 'Evans' 10%), with mean and SD among the five cultivars of 3.8 and 4.1, respectively (Supplemental Table S2a). For the 31 cultivars that had been evaluated for DUS, the range, mean percentages, and SD of SNP heterogeneity were 0–5, 1.8, and 1.3%, respectively (Supplemental Table S2b).

3.3 | Pairwise single nucleotide polymorphism similarity

Pairwise genetic similarities (Supplemental Table S3) among the 322 cultivars (also identified in Supplemental Table S1) ranged from 44 to 100%, distributed as a bell-shaped curve with a mean of 64.0% and standard deviation of 6.7% (Figure 3a). The upper 1% of this distribution ranged from 79 to 100% similarity. Distribution of pairwise similarities among members of cultivar pairs for the subset of 187 plant variety protected varieties ranged from 52.9 to 99.5% with a mean of 67.1% and standard deviation of 6.1% (Figure 3b).



FIGURE 2 Coefficient of variation (CV) curves for different scenarios consistent with the results of the experiments that were conducted

3.4 | Cultivar comparison: single nucleotide polymorphism, pedigree, and morphology

Individual dendrograms showing associations among cultivars using pedigree kinship data (left vertical) and SNP genetic similarities (right vertical) are aligned using a tanglegram (Supplemental Figure S1). Along the lefthand vertical pedigree-kinship side of the tanglegram (Supplemental Figure S1), short branches indicated a high degree of pedigree similarity. For example, cultivars Century and Century 84 together formed a very short branch on the kinship dendrogram because Century 84 was derived following four generations of backcrossing to Century, which resulted a high level of kinship. Along the right-hand vertical SNP side of the dendrogram, these two varieties were also joined on a short branch, which thereby indicated their high genetic similarity. In both dendrograms, higher values along the scale shown at the bottom of Supplemental Figure S1 indicated greater similarity. The diagonal lines between the two dendrograms link the individual leaves for the same cultivars. In the vast majority of cases, the shortest branches on the kinship dendrogram corresponded to the shortest branches on the SNP dendrogram. Similarly, unrelated cultivars in both SNP and kinship dendrograms were positioned with long branches. For example, the kinship dendrogram branches for 'StrainNo18' and 'Kingwa' were completely separated with almost no similarity. In the SNP dendrogram, the branches for these two varieties merged together on the far-right side, which indicated very low genetic similarity.

In contrast, according to known pedigrees, cultivars Edison and Flyer are 25% related by pedigree-kinship but appear genetically more similar (85%) according to a comparison of their SNP profiles. Interestingly, according to SNP comparisons (right-hand vertical of Supplemental Figure S1), both cultivars Flyer and Edison are closely associated with cultivar A3127. Such a close association with A3127 is expected based on the pedigree of Flyer, which includes kinship with A3127 and cultivar Williams 82. However, the only pedigree kinship connection between Flyer and Edison is through Williams 82 as a great grandparent of Edison. These data suggest that either Edison retained much more germplasm originally inherited via Williams 82, an error in its pedigree, or a seed mislabeling error. In summary, tanglegram analysis highlighted that structuration among soybean cultivars according to analyses using SNP data was associated and concordant with known pedigrees. Also, comparisons of associations among cultivars on the basis of kinship as expected from pedigrees with associations based upon genetic similarities directly measured using SNPs provides means to identify possible errors either in recorded pedigree or in the cultivar names ascribed to specific accessions of seed.

Correlation analysis also revealed agreement between SNP-based and pedigree-based kinship similarities amongst the 322 cultivars, r = .77 (Figure 4a), despite the fact that the kinship matrix was sparse, having many



FIGURE 3 Distribution of genetic similarities between (a) each pair of 322 soybean cultivars and for (b) each pair of the 187 plant variety protection cultivar subset (cultivars are identified in Supplemental Table S1

zero or missing kinship values, thereby leading to the possible underestimation of true kinship. There are some notable outliers, further scrutiny of which provides useful information. For example, there were two pairs each with approx. 50% SNP similarity between members but with pedigree kinship similarities of approximately 75 and 100%, respectively. Both these pairs included the cultivar Kingwa as one member with the other being cultivar Peking. The cultivar Peking is a landrace introduced from China with a pedigree and source labelling in GRIN of Beijing, China, 1906. However, there are four accessions of Peking with SNP data in GRIN representing accessions



FIGURE 4 (a) Scatter plot of pairwise cultivars comparing SNPbased and pedigree-based kinship similarities for 322 soybean cultivars. (b) Scatter plot of pairwise cultivars comprised of members comparing SNP-based and pedigree-based kinship similarities for kinship values >0.25–1.0 or 25–100% similar from the set of 322 soybean cultivars

donated in 1954, 1964, and two in 1979. The cultivar Kingwa was selected from Peking in 1921. The different placement of these cultivar pairs therefore reflects different SNP profiles for accessions labelled Peking. Different biotypes can be expected to occur as a result of continued further selfing of the original landrace material. An opposite example, where percentage SNP similarity is far greater than would be anticipated on the basis of known pedigree also occurs, for example by a cultivar pair with 98% SNP similarity yet only 24% similar on the basis of known pedigree. Three explanations for this association of cultivars include the following: (a) the result of selection toward one of the breeding parents, (b) mislabeling of pedigree, and (c) mislabeling of seed. However, for the purposes of this study, it is most appropriate to compare cultivars that are more, rather than less, related and which have relatively high degrees of SNP similarity. Consequently, we also presented comparisons of SNP and pedigree similarities for those pairs of varieties with a greater depth of pedigree data (>0.25 CP) and with SNP similarities >89% (Figure 4b). Here the correlation was reduced (r =.63) with ~ 96% SNP similarity for the point on the linear regression line at 100% kinship (Figure 4b).

Cultivar pairs with SNP similarity >90% can be classified as (a) very highly related by more than four backcrosses of the recurrent parent; (b) lesser degrees of relatedness including reselections from the same cultivar, three or fewer backcrosses of the recurrent parent, full-sibs, halfsibs, and 50% common parentage; and (c) lesser or unrelated (Table 3). For cultivars with >97% SNP similarity, all but a single cultivar pair ['A.K. (Harrow)'-'Illini'] were the result of multiple backcrosses to introduce either racespecific resistances to Phytophthora or, for a single pair ('Williams'-'Kunitz') to remove the trypsin inhibitor gene (Bernard, Hymowitz, & Cremeens, 1991). Cultivars A.K. (Harrow) and Illini were selections from the same source A.K. (Supplemental Table S1). Cultivars within the range of 95-96.9% SNP similarity represented a mix of related cultivars with a predominance of highly related cultivars, likewise reflecting breeding practice to introduce race-specific Phytophthora resistance. Cultivars within the range 90-94.9% SNP similarity represented a mix of related and unrelated cultivars with a predominance of unrelated pairs when SNP similarities fell below 94%. Cultivar pairs with similarity >90% that differed for nondisease morphological characteristics are, with SNP similarity in parentheses: 'Cutler' and 'Cutler 71' (97.3%) differed for plant height; 'SRF' and 'Clark' (97.0%) differed for leaf shape, seeds per pod, grams per 100 seeds; 'S1492' and 'B216' (96.5%) differed for maturity and plant type; 'Camp' and 'Vance' (96.5%) differed for seed size; 'Wayne' and 'SRF307B' (96.4%) differed for leaf shape, seed size, number seeds per pod, hilum color; 'Century' and 'Century 84' (95.5%) differed for plant height; 'Resnik' and 'Flyer' (94.7%) differed for maturity; 'Corsoy' and 'Hardin' (94.1%) differed for maturity; 'A3127' and 'Flyer' (94%) differed for maturity; 'GR8836' and 'Flyer' (93.1%) differed for maturity; 'Bedford' and 'Forrest' (92.6%) differed for maturity; and 'Vertex' and 'Sandusky' (91.5%) differed for maturity and pod color.

There is good agreement between pedigree (kinship) and SNP similarity for the 187 subset of cultivars where degree of pedigree-kinship relatedness rises as genetic similarities between members of each pair also rise according to comparisons of their SNP profiles (Figure 5). Scatter plots of cultivar pairs are shown for each of four ranges of pedigree-kinship (0–100%, 25–100%, 5–100%, and 75–100%) (Figures 5a–5d, respectively). Correlations between pedigree-kinship and SNP similarities for cultivar pairs

SNP range percentage	No. cultivar pairs	High		Interm	ediate ^b	Unrelated		
similarity	in each SNP class	NO.	%	No.	%	No.	%	
99–100	4	3	75	1	33	0	0	
98–98.9	4	4	100	0	0	0	0	
97–97.9	11	11	100	0	0	0	0	
96–96.9	8	5	62.5	3	37.5	0	0	
95-95.9	11	7	63.6	3	27.3	1	9	
94–94.9	5	1	20	4	80	0	0	
93–93.9	3	0	0	2	66.6	1	33.3	
92-92.9	2	0	0	2	100	0	0	
91–91.9	9	1	11	6	66.6	2	22.2	
90-90.9	15	0	0	8	53.3	7	46.7	

TABLE 3 Summary of parental pedigree backgrounds for cultivars >89.9% similar according to comparisons of single nucleotide polymorphism (SNP) profiles from the set of 322 cultivars' pedigree relatedness categories of parents

^aMore than four backcross generations.

^bIncludes less than four backcross generations, full-sibs, half-sibs, and 50% common parentage.



FIGURE 5 Scatter plots of pairwise cultivars comprised of members >89.9% SNP similarity comparing SNP-based and pedigree-based kinship similarities for cultivars with percentage kinship similarity values in the range (a) 0–100%, (b) 25–100%, (c) 50–100%, and (d) 75–100%

ranged from r = .46 to r = .84. Highest correlations were found when considering the entire pedigree–kinship range (r = .66), or when only cultivar pairs within the highest percentage pedigree–kinship range of 75–100% were included in the comparison with SNP-base similarities (r = .84). The former comparison covers the widest range of pedigree– kinship values while the latter kinship range involves cultivar pairs comprised of members with individually the greatest depth of pedigree data vs. other cultivars.

Euclidean and simple percentage morphological distances, SNP percentage similarities, and percentage similarity pedigree–kinship data, for 42 cultivar pairs with



FIGURE 6 Scatter plot of cultivar pairs for percentage SNP similarity (>89.9%) and percentage difference in expressed morphological characteristics from the subset of 187 cultivars that met Distinct Uniform Stable (DUS) eligibility requirements plus additional USDA Plant Variety Protection (PVP) Office reference cultivars. Three cultivar pairs with the least differences for expression of morphological characteristics are highlighted within an ovoid. Additional pedigree, kinship, PVP status, and intercultivar distances on the basis of morphological and physiological characteristics are presented in Supplemental Table. S3

members >89.9% SNP similarity are presented in Supplemental Table S4, using reports of their morphological and physiological characteristics that were provided by the U.S. PVP Office. These cultivar pairs were drawn from the subset of 187 PVP cultivars, which was itself a subset of the 322 cultivars. Occasionally, these pairs included cultivars that were not themselves PVP, but which are included in the U.S. PVP reference set of cultivars of common knowledge for determination of distinctness. Data reported in column N of Supplemental Table S4 indicates the PVP status of each cultivar.

Figure 6 presents a scatterplot of percentage SNP similarities between pairs of cultivars with >89.9% SNP similarity using Euclidean distances calculated using morphological (but excluding physiological) data provided by the U.S. PVP Office with a correlation r = .52. When differences among cultivars for physiological characteristics including race-specific disease resistance, trypsin inhibitor, and seed protein composition were also included, the correlation dropped markedly to 0 (data not shown). This drop in correlation between SNP similarity and overall morphological and physiological similarity is expected as a result of the introduction of different physiological characteristics from donor cultivars while subsequently retaining high genetic conformity with the recipient cultivar following multiple generations of backcrossing using that cultivar.

Three pairs of cultivars highlighted with an ovoid in Figure 6 are particularly informative because they express the least differences for comparisons of morphological characteristics. First, cultivars Wells and Wells II had a SNP similarity of 99.54% and were morphologically indistinguishable (Supplemental Table S4; Figure 6). However, these cultivars additionally express different physiological reactions to *Phytopthora* spp. (Supplemental Table S4) (Wilcox, Athow, LLaviolette, Abney, & Richards, 1979). Second, cultivars S1492 and B216 were the least morphologically different (96.51% SNP similarity, 0.69 Euclidean distance). Third, cultivars Kunitz and Regal were the next most different morphologically (95.51% SNP similarity, 1.175 Euclidean distance).

Euclidean distance data result from a synthesis of data described in multidimensional space and combining information from all characteristics (Sneath & Sokal, 1962). Consequently, it is also informative to examine differences between cultivars on a simpler basis. The number and percentage of morphological (excepting physiological) characteristics that differed between cultivars (Supplemental Table S4, columns H and I, respectively), ranged from 0 to 7 (50%). The cumulative percentage of 42 cultivar pairs that expressed morphological differences (Supplemental Table S4, column H) were 0% (>99% SNP similarity), 10% (>98% SNP similarity), 21% (>97% SNP similarity), and 38% (>96% SNP similarity). In other words, 38% of these cultivar pairs were comprised of members expressing different morphologies with SNP similarities ranging from 96 to 98.6% (Supplemental Table S4). Of cultivars <96% similar by SNPs, members of all but a single pair, 'Vinton' and 'Vinton 81' (94.55% SNP similarity, one morphological difference) differed by three or more morphological characteristics. Consequently, a SNP similarity of 96% between soybean cultivars represents a conservative point of demarcation between cultivars that have morphological differences and those that do not.

3.5 | Single nucleotide polymorphism set robustness and lab concordance

Mantel test correlations for pairwise genetic similarity among 322 soybean varieties for the 5346 SNPs and several SNP subsets were very high (>.95). When SNP subsets were reduced to 668 SNPs or fewer, with the introduction of up to 2.5% mistyped data, correlations remained relatively high and robust, though in some cases, dropping to \sim .70 (Supplemental Table S5 and summarized in Table 4). Subsets of SNPs that were selected to maintain even genome coverage with a constant level of expected heterozygosity had a slightly higher level of robustness as

TABLE 4 Summary of Mantel test correlations for pairwise distances among 322 soybean cultivars for the 5346-single nucleotide polymorphism (SNP) set and each subset; full data are presented in Supplemental Table S5

SNP set size		5346	5346	5346	2673	2673	2673	1336	1336	1336	668	668	668	334	334	334	167	167	167
	Percentage mistype	0	1	2.5	0	1	2.5	0	1	2.5	0	1	2.5	0	1	2.5	0	1	2.5
5346	0	-	1.000	.999	.981	.980	.979	.973	.971	.969	.964	.961	.955	.937	.933	.923	.905	.897	.885
5346	1	1.000	-	.999	.980	.980	.979	.972	.971	.969	.964	.960	.955	.937	.932	.923	.904	.897	.885
5346	2.5	.999	.999	-	.979	.979	.978	.972	.970	.968	.963	.959	.954	.936	.931	.922	.903	.896	.883
2673	0	.996	.995	.994	-	.999	.998	.992	.991	.988	.980	.977	.971	.955	.950	.940	.913	.905	.892
2673	1	.995	.994	.994	.999	-	.999	.992	.990	.988	.980	.976	.970	.954	.949	.939	.912	.904	.892
2673	2.5	.993	.993	.992	.997	.998	-	.990	.989	.986	.979	.975	.969	.953	.948	.938	.911	.903	.891
1336	0	.986	.986	.985	.982	.981	.979	-	.998	.996	.985	.982	.976	.959	.955	.945	.913	.905	.892
1336	1	.984	.984	.983	.980	.979	.977	.998	-	.997	.983	.980	.975	.957	.952	.942	.911	.903	.890
1336	2.5	.981	.980	.980	.977	.976	.975	.995	.997	-	.980	.977	.972	.954	.949	.939	.909	.900	.887
668	0	.971	.971	.969	.968	.967	.965	.959	.957	.955	-	.997	.992	.968	.964	.954	.925	.916	.903
668	1	.968	.967	.966	.965	.964	.962	.956	.954	.951	.997	-	.995	.965	.961	.951	.923	.914	.900
668	2.5	.963	.962	.961	.960	.959	.957	.951	.949	.946	.992	.995	-	.961	.956	.946	.918	.909	.895
334	0	.945	.944	.944	.939	.938	.936	.929	.927	.924	.916	.914	.909	-	.994	.985	.946	.937	.922
334	1	.940	.939	.939	.934	.933	.931	.924	.922	.919	.911	.909	.904	.994	-	.991	.941	.932	.918
334	2.5	.931	.931	.931	.926	.925	.923	.916	.914	.911	.903	.900	.895	.985	.990	-	.931	.922	.907
167	0	.881	.881	.879	.878	.877	.874	.865	.862	.859	.848	.846	.841	.841	.838	.829	-	.991	.975
167	1	.873	.873	.871	.870	.868	.866	.857	.855	.852	.840	.838	.832	.834	.831	.823	.990	-	.984
167	2.5	.858	.858	.857	.855	.854	.851	.843	.840	.837	.826	.824	.819	.819	.817	.809	.974	.984	-

TABLE 5Ranges of concordances between different laboratories for 5103 single nucleotide polymorphism profiles of the same seedsfrom each of 34 different soybean cultivars (Supplemental Table S1). Numbers below the diagonal relate to the profiling by the laboratories ofthe first seed extract, numbers above the diagonal refer to the second DNA extract for each cultivar

Laboratories	Bayer-seed 2	Dow-seed 2	Eurofins-seed 2	Gene Seek-seed 2	Pioneer-seed 2
Bayer-seed 1	-	.9987–.9998	.99819997	.9888–.9998	.9981–.9997
Dow-seed 1	.9984–.9998	-	.9979–1.0	.9894–1.0	.9988-1.0
Eurofins-seed 1	.99859998	.9992-1.0	-	.9889–1.0	.9987–1.0
Gene Seek-seed 1	.99889998	.9984–1.0	.9994–1.0	-	.9985–1.0
Pioneer-seed 1	.99849998	.9984-1.0	.9997–1.0	.9994–1.0	-

evidenced by levels of correlation exceeding 0.95 vs. SNP subsets that were selected randomly.

Levels of concordance between genotyping scores generated by each laboratory using the same DNA were very high (>.9888) (Table 5). Given each laboratory used their regular methodology in all aspects of SNP analysis (see also methods), then any variables associated with lab processes, including allele calling, had very minimal effects on SNP data that were generated and reported.

3.6 | Chronological monitoring of genetic diversity

During the period of 30 yr when these cultivars were developed (Supplemental Table S1) the means and upper bound of pedigree–kinship-based similarities between parents of breeding populations rose slightly from 18 to 25% and from 56 to 60%, respectively. Similarly, means and upper bounds for levels of SNP similarities between these same parents and during this period also rose slightly from 65 to 69% and from 83 to 86%, respectively.

4 | DISCUSSION

The determination of cultivar distinctness and its counter state, cultivar sameness or identification, uses the principles of numerical taxonomy (Moss & Hendrickson, 1973), extended below the level of species, to the level of cultivar. The list of characteristics used to describe and to compare cultivars inevitably represents a restricted set of data because it is impossible to "obtain every conceivable shred of data" (Moss & Hendrickson, 1973). For example, agronomic performance data are impractical to use for DUS evaluation because they are very influenced by G×E interactions and require many more resources, especially field space and time to obtain than morphological data. However, much, if not most, of the morphological characteristics used to determine DUS in soybean are also subject to G×E effects and correlations among characteristics, thereby undermining their suitability for application in taxonomic analysis (Sneath & Sokal, 1962). In contrast, while it was the case during previous decades that genotypic data were not directly available for comparisons among organisms (Moss & Hendrickson, 1973), including among cultivars, this deficiency is demonstrably no longer the case.

The increasing scale of DUS testing conducted with primary, if not complete, reliance on comparisons of morphological characteristics threatens to undermine abilities to efficiently and effectively provide PVP for new soybean cultivars because of the numerous challenges noted previously. As a result, "it is almost impossible to have and maintain a full overview of [varieties of] common knowledge. The rapid development of new varieties as a result of intensive molecular assisted breeding and increased global character of the plant breeding industry, makes it an already hard and soon impossible task to keep track of [varieties of] common knowledge in living form in seeds or plants." (van Ettekoven, 2017). Wallace (2017) noted that the growth in reference collections is making DUS systems "difficult to manage ... resulting in a testing system that is becoming unsustainable."

Molecular marker data provide opportunities to facilitate the DUS process on a national, regional, and global basis because of their immunity from G×E effects, public availability, cost-effectiveness, and robustness (De Riek, 2001). Establishing a specific set of SNP loci that are publicly available creates a level playing field for all applicants and prevents biased sampling or "cherry-picking" of SNP loci to suit short-term goals of specific applicants. Single nucleotide polymorphism data provide a far more repeatable, efficient, and cost-effective means of characterizing soybean cultivars because of the absence of G×E effects and minimal genotype \times laboratory effects in contrast to the time, field, and personnel resources required to record and to compare the expression of morphological characteristics. Furthermore, use of a single set of SNPs can contribute not only to national or regional harmonization but also to global harmonization.

There are several means to assay SNPs, including to generate sequence data, and additional platforms to acquire SNP data can be expected to be developed in the future. We chose to use an array platform that is publicly available that allows many SNPs to be assayed simultaneously through multiplexing. Public availability is a prerequisite to allow all interested parties, including PVP agencies and breeders, to have equal access to use SNPs fully within their respective programs. However, this study is not intended as an endorsement of any specific technological platform to inquire SNP data. Nonetheless, we recognize that associations among cultivars can be dependent upon SNP number, degree of map coverage, abilities of different laboratories to repeat results, and ascertainment bias. Consequently, we examined the effects of using subsets of SNPs and the robustness of results in the face of missing data and as generated in five different laboratories. Robustness was high in the face of both missing and mistyped data. Levels of concordance as a result of SNP profiling, quality control, scoring, and reporting of SNP data among five different laboratories was very high (>.99) (Table 5).

Ascertainment bias can result from the selection of highly discriminating characteristics using one set of germplasm but which might then be found to be less usefully discriminating among another, usually unrelated, set of germplasm. For example, while the BARCSoySNP23 selected by Yoon et al. (2007) was able to uniquely identify 132 soybean cultivars, including 36 U.S. cultivars that "were seemingly identical based upon maturity, seed coat color, hilum color, cotyledon color, leaflet shape, flower color, pod color, pubescence color and plant habit;" this set of SNPs was predicated solely on their collective ability to discriminate among those specific soybean cultivars. In contrast, the selection of the BARCSoySNP6K was predicated upon successive evaluations of discrimination involving a very broad base of soybean germplasm. The initial SoySNP50K selection was purposely made using a very diverse set of soybean germplasm, including 96 diverse landraces collectively from three countries, 96 elite cultivars of soybean from North America released by public sector breeding programs from 1990-2000, and 96 wild soybean accessions collectively from four countries (Song et al., 2013). Song et al. (2014) described the selection of SNPs from those that are present in the BARCSoy50K with the goal to still capture as much haplotype diversity as possible. Other important selection criteria included MAF, the quality of genotyping data, even genomic spacing, and representative of both euchromatic and heterochromatic regions of the genome. Song et al. (2014) concluded that "the BARCSoySNP6K beadchip will be an excellent tool for the detection of quantitative trait loci and for assessing genetic diversity." In the latter regard, Liu et al. (2017b) found that associations among 577 Chinese and U.S. soybean cultivars using the SoySNP6K reflected the geographical origins and pedigrees of the cultivars, thereby showing no indication of ascertainment bias within or among these sets of soybean germplasm. Consequently, the suitability of other platforms to provide equivalent results as those presented here should only require demonstration of their equivalency in repeatably reporting SNP data.

4.1 | Establishing a distinctness threshold

4.1.1 | Relevant factors to be considered in order to maintain the current level of intellectual property protection

Regardless of data source, whether it be morphological, physiological, or molecular markers, determining a distinctness threshold leads to the fundamental question of how to define minimum distance or, in other words, "how different is different?" (Wallace, 2017). We concur that the introduction of more efficient testing must take into account the current level of IPP resulting from the grant of PVP as a result of the comparison of morphologically expressed characteristics (De Riek, 2001). In this regard, use of SNP data in the context of determining distinctness has been critiqued because, in the extreme case, distinctness could be determined on the basis of a single SNP difference. However, morphological or physiological differences also can be dependent upon single-gene and even single-SNP differences (Liu et al., 2010; Yan et al., 2014). In the event of concerns about distinctness being determined by a single-gene difference, authorities can introduce a greater threshold requirement of difference in the expression of morphological or physiological characteristics, such as that practiced by GEVES, the French PVP testing authority using the GAIA, or weighted characteristic approach (Gregoire, 2003; 2007; Maton et al., 2014; Thomasset et al., 2015; UPOV, 2010). Similarly, with regard to the use of SNPs, the possibility of distinctness being dependent upon either a single or small number of base pairs, which could thereby undermine an effective level of IPP in the context of PVP, is removed by establishment of a SNP percentage similarity threshold. Consequently, we took an approach that sought to recalibrate the current approach using the comparative expression of morphological characteristics to an equivalent approach using SNP data, thereby maintaining the current level of IPP provided by PVP.

4.1.2 | Observations contributing to calibration of a single nucleotide polymorphism-based distinctness threshold

Bulk samples of 10–15 individual plants per cultivar were found to provide a basis for generating DNA samples that are representative of each cultivar. We then sought a SNP percentage similarity that could provide an equivalent determination of distinctness as have comparisons of expressed morphological characteristics. We initially compared SNP-based similarities and pedigree-based kinships among 322 soybean cultivars. We also included comparisons of differences in expression of morphological and physiological characteristics with information gleaned from most closely similar cultivar notes published in PVP certificates for cultivar pairs from this set where members were >89.9% similar according to their comparative SNP profiles. This set of cultivars included those that had been declared as DUS for the purposes of obtaining PVPs and many other cultivars developed in the public domain that had not been submitted for PVP certification (Supplemental Table S1; Table 3; Figure 4). These data indicated a possible SNP threshold range of 93-97% similarity that potentially could be concordant with an evaluation of distinctness.

We then examined in greater detail correlations among SNP and pedigree–kinship data for a subset of 187 PVP cultivars that had been found to meet DUS requirements for PVP certification (Supplemental Table S4; Figure 5). We also examined correlations of differences in morphological and physiological characteristics with SNP similarity for members of pairs with >89.9% SNP similarity using morphological and physiological data provided by the U.S. PVP Office (Supplemental Table S4; Figure 6). With the exception of cultivars Wells and Wells II, all cultivars could be distinguished by their expression of at least one morphological characteristic (Figure 6).

While the initial round of analyses suggested evidence of distinctness in the range of 93-97% SNP similarity, this second round of analysis suggested that 96% SNP similarity could provide a suitable threshold for determining distinctness, albeit one that is possibly conservative given examples of distinctness according to the expression of morphological characteristics for soybean cultivars that were up to 98.6% similar according to SNP data. Consequently, we note that a 96% SNP similarity threshold is perhaps conservative and does not necessarily represent an upper bound for declaring distinctness. Consequently, cultivars that are >96% similar according to SNP data, but which also differ in their morphological or physiological attributes, would still be classed as distinct so long as these characteristics are the ultimate test of distinctness. The 96% similarity threshold as an initial evaluation of distinctness was independently validated by several U.S. soybean breeding companies that are active in submitting applications for soybean to the U.S. PVP Office. They examined SNP data for soybean cultivars that were either recently developed or under development. They reported validation of this threshold (S. Schnebly, personal communication, 2018; Y. Bin, personal communication, 2019; T. Hamilton, personal communication, 2020). Robustness of SNP profiling reported from five different laboratories was very high (Table 5).

4.2 | Uniformity

It is well understood from an elementary knowledge of Mendelian genetics that application of a typical breeding scheme for soybean (Diwan & Cregan, 1997), whereby two parental genotypes are hybridized to produce an F_1 population which is then "followed by several rounds of singleseed descent via self-mating and subsequent seed increase generations" (Haun et al., 2011), inevitably results in a certain percentage of segregating loci, which then become fixed for alternate alleles. The process of conducting successive cycles of self-pollination results in the presence of slightly different genetic strains, which appear as heterozygous SNP loci when profiled using bulk samples of plants of an individual cultivar.

Residual heterogeneity can be retained not only for SNPs but also for loci affecting the expression of morphological and agronomically important characteristics including those associated with responses to stress (Espinosa et al., 2015). For example, residual variation within soybean cultivars Benning, Cook, and Haskell, each of which appeared uniform when grown according to common agronomic practice, was sufficient to allow up to seven new morphologically and agronomically distinct cultivars to be selected from individual plants when planting densities were much reduced (Fasoula & Boerma, 2005; 2007; Fasoula et al., 2007a; 2007b; 2007c; Haun et al., 2011; Varala, Swaminathan, Li, & Hudson, 2011; Yates, Boerma, & Fasoula, 2012). Genetic heterogeneity can also result from mutation, intragenic recombination, unequal crossing over, DNA methylation, excision or insertion of transposable elements, and gene duplication (Cullis, 1990; Kidwell & Lisch, 2002; Morgante et al., 2005; Rasmusson & Phillips, 1997; Sandhu et al., 2017).

Concerns have been expressed that use of marker data in DUS evaluation might lead to the introduction of unrealistically and unnecessarily high levels of uniformity being required at the DNA sequence level, leading to higher resource demands during breeding and seed multiplication (International Seed Federation, 2012). In this respect, we are unaware of reports suggesting that a reliance upon comparisons of morphologically expressed characteristics to establish uniformity to standards required by PVP offices has been inadequate or unsatisfactory to support the stable reproduction of soybean cultivars. We therefore chose to determine a SNP threshold in respect of uniformity through calibration informed by measuring the degree of intracultivar heterogeneity of SNP loci of cultivars that had already been declared to have met DUS criteria (Supplemental Table S2b). Levels of intracultivar SNP heterogeneity were low (range 0-5%, mean 1.8%, standard deviation 1.3%) for 35 commercially developed varieties. Soybean cultivars that were the least similar on the basis of their SNP profiles exhibited 46-55% SNP similarity, with the majority being less than 62–69% similar by SNPs (Figure 3). Consequently, these levels of SNP heterogeneity are consistent with a commonly used breeding strategy of bulking individual plants at, or close to, the F₄ stage of inbreeding. With regard to uniformity, a percentage homozygosity threshold approach using marker data has likewise been proposed as a substitute for fieldbased studies of uniformity in wheat (Triticum aestivum L.) (Wang et al., 2014).

4.3 | Chronological monitoring of genetic diversity

Comparisons of pedigree–kinship and SNP-based genetic similarities between pairs of soybean cultivars used to develop F_2 segregating populations for further crossing and selection did not provide much evidence for a narrowing of the soybean germplasm base, at least for the purpose of creating those populations and during the three-decadal period 1970–1989. These results support that challenges to establish distinctness for soybean cultivars derive primarily, if not entirely, from an inherent relative lack of their distinguishing power in domesticated soybean.

5 | CONCLUDING COMMENTS

In conclusion, the analytical approach we have described is similar to those previously reported and which have contributed to a Model 2 approach involving management of reference collection, including procedures that are routinely implemented for DUS examinations of maize inbred lines in France (Maton et al., 2014; Thomasset et al., 2015). Nonetheless, the approach reported here differs by its analytical basis being comprised of soybean cultivars released and evaluated during a three-decadal period with each cultivar having met DUS eligibility requirements based on morphological characteristics and thereby each having been qualified for and granted PVP. With respect to a similar application of molecular marker data, Song et al. (2015) noted that "because a limited number of agronomic or morphological traits are available..., profiling each accession in the USDA Soybean Germplasm Collection with a large number of molecular markers is essential to understand the level of repetitiveness, thus increasing the efficiency of germplasm preservation, characterization, and promoting the more efficient utilization of the genetic resources in soybean breeding programs." This description of the application of SNP data in the field of genetic resource conservation reflects a similar need and application to establish the criterion of distinctness for the granting of plant breeders' rights. Ultimately, we conclude that the methodology of usage of molecular data provided here meets the criteria of (a) maintains existing levels of IPP (De Riek, 2001), (b) provides a level playing field for all breeders regardless of their resource capabilities, (c) makes the process more efficient and potentially more harmonized globally, (d) does not add costs and may reduce costs of conducting DUS testing for applicants and PVP agencies, and (e) does not require levels of uniformity that are unrealistic, overly expensive, unnecessary, or impractical to achieve.

ACKNOWLEDGEMENTS

We wish to thank the American Seed Trade Association for their support and the U.S. PVP Office and USDA GRIN system for the provision of morphological data and for the public availability of soybean cultivars bred in the public domain or that were developed by the commercial sector and made publicly available following expiration of their PVP status. We acknowledge the expertise of Dr. Kevin Wright in generating and providing an explanatory note on the tanglegram analysis. We thank all the persons involved in the five laboratories involved in generating and scoring SNP data. We thank the U.S. PVP Office for the provision of morphological and physiological data in electronic format from public soybean PVP records.

ORCID

Mark A. Mikel https://orcid.org/0000-0001-5364-0907 *J.S.C. Smith* https://orcid.org/0000-0001-6828-8205

REFERENCES

- Adams, S. (1996). Sorting look-alike soybeans: Genetic fingerprinting aids plant variety protection. *Agricultural Research*, 44, 12– 13. Retrieved from https://agresearchmag.ars.usda.gov/1996/aug/ soy/
- Akond, M., Liu, S., Schoener, L., Anderson, J. A., Kantartzi, K. Stella, Meksem, K., ... Kassem, M. (2013). SNP-based genetic linkage map of soybean using the SoySNP6K Illumina Infinium beadchip genotyping array. *Journal Plant Genome Science*, 1, 3. https:// doi.org/10.5147/jpgs.2013.0090
- Bandillo, N. B., Lorenz, A. J., Graef, G. L., Jarquin, D., Hyten, D. L., Nelson, R. L., & Specht, J. E. (2017). Genome-wide association mapping of qualitatively inherited traits in a germplasm collection. *Plant Genome*, 10, 1–18. https://doi. org/10.3835/plantgenome2016.06.0054
- Bernard, R. L., Hymowitz, T., & Cremeens, C. R. (1991). Registration of 'Kunitz' soybean. Crop Science, 31, 232–233. https://doi.org/10. 2135/cropsci1991.0011183X003100010059x

- Boldt, A. S., Sediyama, T., Nogueira, A. P. O., Matsuo, E., & Teixeira, R. C. (2007). Influência do tamanho de semente na caracterização de descritores adicionais de soja. In *Reunião de pesquisa de soja da região central do Brasil* (pp. 120–122). Londrina, Brazil: Embrapa Soja.
- Bundessortenamt. (2017). Federal plant variety office: Plant breeders' rights and national listing. Hannover, Germany: Bundessortenamt.
- Campante, P. (2018). A glance at the Brazilian seed market. Retrieved from http://news.agropages.com/News/NewsDetail—26900.htm
- Camussi, A., Spagnoletti Zeuli, P. L., & Melchiorre, P. (1983). Numerical taxonomy of Italian maize populations: Genetic distances on the basis of heterotic effects. *Maydica*, 28, 411–424.
- Cockram, J., Jones, H., Norris, C., & O'Sullivan, D. M. (2012). Evaluation of diagnostic molecular markers for DUS phenotypic assessment in the cereal crop, barley (*Hordeum vulgare ssp. vulgare L.*). *Theoretical and Applied Genetics*, *125*, 1735–1749. https://doi.org/ 10.1007/s00122-012-1950-3.
- Comstock, R. E., & Moll, R. H. (1963). Genotype-environment interactions. In W. D. Hanson & H. F. Robinson (Eds.), *Statistical genetics and plant breeding* (pp. 164–196). Washington DC: National Academy of Sciences–National Research Council.
- Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation For Statistical Computing. Retrieved from https://R-project.org/
- Community Plant Variety Office. (2017). Protocol for tests on distinctness, uniformity and stability. *Glycine max* (L.) Merrill. Soya bean. Retrieved from https://cpvo.europa.eu/sites/default/files/ documents/glycine_max_0.pdf
- Craviotti, C. (2015). MultiLatin agribusiness: The expansion of Argentinian firms in Brazil. Working paper 5. Amsterdam, The Netherlands: BRICS Initiative for Critical Agrarian Studies (BICAS).
- Cullis, C. A. (1990). DNA rearrangements in response to environmental stress. *Advances in Genetics*, *28*, 73–97. https://doi.org/10.1016/ S0065-2660(08)60524-6
- da Silva, A. F., Sediyama, T., Borem, A., da Silva, F. L., dos Santos Silva, F. C., & Bezerra, A. R. G. (2017). Registration and protection of cultivars. In F. L. da Silva, A. Borém, T. Sediyama, & W. H. Ludke. (Eds.), *Soybean breeding* (pp. 427–440). Cham, Switzerland: Springer Nature.
- De Riek, J. (2001). Are molecular markers strengthening plant variety registration and protection? *Acta Horticulturae*, *552*, 215–224. https://doi.org/10.17660/ActaHortic.2001.552.24
- de Vienne, D. M. (2019). Tanglegrams are misleading for visual evaluation of tree congruence. *Molecular Biology and Evolution*, *36*, 174– 176. https://doi.org/10.1093/molbev/msy196
- Dinkins, R. D., Keim, K. R., Farno, L., & Edwards, L. H. (2002). Expression of the narrow leaflet gene for yield and agronomic traits in soybean. *Journal of Heredity*, 93, 346–351. https://doi.org/ 10.1093/jhered/93.5.346
- Diwan, N., & Cregan, P. B. (1997). Automated sizing of fluorescentlabeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. *Theoretical and Applied Genetics*, 95, 723–733. https://doi.org/10.1007/s001220050618
- Dos Santos Silva, F. C., Sediyama, T., da Silva, A. F., Bezerra, A. R. G., Rosa, D. P., Ferreira, L. V., & Cruz, C. D. (2016). Identification of

new descriptors for differentiation of soybean genotypes by Gower algorithm. *African Journal of Agricultural Research*, *11*, 961–966. https://doi.org/10.5897/AJAR2015.10158

- Espinosa, K., Boelter, J., Lolle, S., Hopkins, M., Goggi, S., Palmer, R. G., & Sandhu, D. (2015). Evaluation of spontaneous generation of allelic variation in soybean in response to sexual hybridization and stress. *Canadian Journal of Plant Science*, 95, 405–415. https://doi. org/10.4141/cjps-2014-324
- Fang, C., Y, Ma, Wu, S., Liu, Z., Wang, Z., Yang, R., ... Tian, Z. (2017). Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biology*, *18*, 161. https://doi.org/10.1186/s13059-017-1289-9
- Fasoula, V. A., & Boerma, H. R. (2005). Divergent selection at ultralow planting density for seed protein and oil content within soybean cultivars. *Field Crops Res*, 91, 217–229. https://doi.org/10.1016/ j.fcr.2004.07.018
- Fasoula, V. A., & Boerma, H. R. (2007). Intra-cultivar variation for seed weight and other agronomic traits within three elite soybean cultivars. *Crop Science*, 47, 367–373. https://doi.org/10.2135/ cropsci2005.09.0334
- Fasoula, V. A., Boerma, H. R., Yates, J. L., Walker, D. R., Finnerty, S. L., Rowan, G. B., & Wood, E. D. (2007a). Registration of five soybean germplasm lines selected within the cultivar 'Benning' differing in seed and agronomic traits. *Journal of Plant Registrations*, *1*, 156–157. https://doi.org/10.3198/jpr2006.03.0198crg
- Fasoula, V. A., Boerma, H. R., Yates, J. L., Walker, D. R., Finnerty, S. L., Rowan, G. B., & Wood, E. D. (2007b). Registration of six soybean germplasm lines selected within the cultivar 'Haskell' differing in seed and agronomic traits. *Journal of Plant Registrations*, 1, 160–161. https://doi.org/10.3198/jpr2006.03.0200crg
- Fasoula, V. A., Boerma, H. R., Yates, J. L., Walker, D. R., Finnerty, S. L., Rowan, G. B., & Wood, E. D. (2007c). Registration of seven soybean germplasm lines selected within the cultivar 'Cook' differing in seed and agronomic traits. *Journal of Plant Registrations*, *1*, 158–159. https://doi.org/10.3198/jpr2006.03.0199crg
- Fehr, W. R. (1978). Breeding. In A. G. Norman (Ed.), *Soybean: Physiology, agronomy, and utilization* (pp. 119–155). New York: Academic Press.
- Galili, T. (2015). Dendextend: An R package for visualizing, adjusting, and comparing trees of hierarchical clustering. *Bioinformatics*, *31*, 3718–3720. https://doi.org/10.1093/bioinfo
- Giancola, S. M. Poltri, Lacaze, P., & Hopp, H. E. (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. https://doi.org/10.1023/A:1019923923805
- Gibson, D. (2015). Genetic and physiologic analyses of soybean grain yields in water limited environments (M.S. thesis, Iowa State University). Retrieved from https://lib.dr.iastate.edu/etd/14366/
- Gopal, J., Pandey, S. K., Kumar, V., Kumar, R., Pande, P. C., & Singh, S. V. (2018). Morphological descriptors for DUS testing of potato varieties. *PGR Newsletter*, 154, 40–47. Retrieved from https://www.bioversityinternational.org/fileadmin/PGR/articleissue_154-art_7-lang_en.html
- Gregoire, S. (2003). GAIA software: Crop expert phenotypic distances between varieties. TWC/21/4. Technical working party on automation and computer programs, 21st session Tjele, Denmark, June 10– 13. Geneva, Switzerland: International Union for the Protection of New Varieties of Plants.
- Gregoire, S. (2007). GAIA software: TWC/25/13. Technical working party on automation and computer programs, 25th session Sibiu,

Crop Science

21

Romania, September 3–6, 2007. Geneva, Switzerland: International Union for the Protection of New Varieties of Plants.

- Hariprasanna, K. (2018). Distinctness, uniformity and stability testing in sorghum. In C. Aruna, K. B. R. S. Visarada, B. V. Bhat, & V. A. Tonapi (Eds.), *Breeding sorghum for diverse end uses* (pp. 341–366). Cambridge, UK: Woodhead Publishing.
- Haun, W. J., Hyten, D. L., Xu, W. W., Gerhardt, D. J., Albert, T. J., Richmond, T., ... Stupar, R. M. (2011). The composition and origins of genomic variation among individuals of the soybean reference cultivar Williams 82. *Plant Physiology*, *155*, 645–655. https:// doi.org/10.1104/pp.110.166736
- Heckenberger, M., Bohn, M., Klein, D., & Melchinger, A. E. (2005). Identification of essentially derived varieties obtained from biparental crosses of homozygous lines: II. Morphological distances and heterosis in comparison with simple sequence repeat and amplified fragment length polymorphism data in maize. *Crop Science*, 45, 1132–1140. https://doi.org/10.2135/cropsci2004.0111
- Huang, L., Zeng, A., Chen, P., Wu, C., Wang, D., & Wen, Z. (2018). Genome wide association analysis of salt tolerance in soybean [*Glycine max* (L.) Merr.]. *Plant Breeding*, 137, 714–720. https://doi. org/10.1111/pbr.12623
- Illumina, Inc. (2015). Agrigenomics, commercial, collaborative, and consortia Infinium arrays. San Diego, CA: Illumina, Inc. Retrieved from https://www.illumina.com/content/dam/illumina-market ing/documents/products/product_information_sheets/infiniumarrays-summary-agrigenomics-1370-2015-008.pdf
- Illumina, Inc. (2016). GenomeStudio® GenotypingModule v2.0 software guide. San Diego, CA: Illumina. Retrieved from https://support.illumina.com/content/dam/illumina-support/do cuments/documentation/software_documentation/genomes tudio/genomestudio-2-0/genomestudio-genotyping-module-v2user-guide-11319113-01.pdf
- International Seed Federation. (2012). *ISF view on intellectual property*. Nyon, Switzerland: International Seed Federation. Retrieved from https://www.worldseed.org/wp-content/uploads/ 2015/10/View_on_Intellectual_Property_2012.pdf
- Jamali, S. H., Cockram, J., & Hickey, L. T. (2019). Insights into deployment of DNA markers in plant variety protection and registration. *Theoretical and Applied Genetics*, *132*, 1911–1929. https://doi.org/10. 1007/s00122-019-03348-7
- Jarman, R. J., & Hampson, A. G. (1991). The use of a computer management system for testing candidate cereal varieties for distinctness, uniformity and stability and the award of plant breeders' rights. *Plant Varieties and Seeds*, 4, 161–168.
- Jones, H., Jarman, R. J., Austin, L., White, J., & Cooke, R. J. (2003). The management of variety reference collections in distinctness, uniformity and stability testing of wheat. *Euphytica*, *132*, 175–184. https://doi.org/10.1023/A:1024642828705
- Jones, H., Norris, C., Smith, D., Cockram, J., Lee, D., and O'Sullivan, D. M. (2013). Evaluation of the use of high-density snp genotyping to implement UPOV Model 2 for DUS testing in barley. *Theoretical* and Applied Genetics, 126, 901–911. https://doi.org/10.1007/s00122-012-2024-2
- Karivaradaraaju, T. V. (2005). Physiological and biochemical techniques for crop variety identification. In K. Rathinavel, S. Manickam, & M. Sabesh (Eds.), *Training manual on DUS test in cotton* with reference to PPV & FR legislation, 2001 (pp. 76–84). Coimbatore, India: Central Institute For Cotton Research Regional Station.
- Khan, A. Z., Khalil, S. K., & Taj, F. H. (2003). Influence of planting date on planting date and plant density on morphological traits of

determinate and indeterminate soybean cultivars under temperate environments. *Journal of Agronomy*, *2*, 146–152. http://doi.org/10. 3923/ja.2003.146.152

- Kidwell, M. G., & Lisch, D. (2002). Transposable elements as sources of genomic variation. In N. L. Craig, R. Craigie, M. Gellert, & A. M. Lambowitz. (Eds.), *Mobile DNA II* (pp. 59–90). Washington, DC: American Society Microbiology Press.
- Korir, N. K., J.Han, L. Shangguan, Wang, C., Kayesh, E., Zhang, Y., & Fang, J. (2013). Plant variety and cultivar identification: Advances and prospects. *Critical Reviews in Biotechnology*, *33*, 112–125. https://doi.org/10.3109/07388551.2012.675314
- Kumar, N. (2014). DUS characterization on morphological biochemical and molecular basis in *Brassica junc*ea L. Czern & Coss and *Brassica carinata* A. Braun (Doctoral dissertation, G.B. Pant University of Agriculture and Technology). Retrieved from http:// krishikosh.egranth.ac.in/bitstream/1/67379/1/Nitin24.PDF
- Kumar, V., Rani, A., Jha, P., Rawal, R., & Husain, S. M. (2017). Molecular identification of dominant cultivars of soybean using simple sequence repeat markers. *Proceedings of the National Academy of Sciences, India Section B: Biological Sciences*, 87, 647–653. https://doi.org/10.1007/s40011-015-0641-7
- Law, J. R., Anderson, S., Jones, E. S., Nelson, B., Mulaomasonovich, E., Hall, B., & Smith, J. S. C. (2011). Approaches to improve the determination of eligibility for plant variety protection: I evaluation of morphological characteristics. *Maydica*, 56, 113–131.
- Lence, S. H., Hayes, D. J., Alston, J. M., & Smith, J. S. C. (2015). Intellectual property in plant breeding: Comparing different levels and forms of protection, *European Review of Agricultural Economics*, 43, 1–29. https://doi.org/10.1093/erae/jbv007
- Liu, B., Watanabe, S., Uchiyama, T., Kong, F., Kanazawa, A., Xia, Z., ... Abe, J. (2010). The soybean stem growth habit gene *Dt1* Is an ortholog of *Arabidopsis TERMINAL FLOWER1*. *Plant Physiology*, *153*, 198–210. https://doi.org/10.1104/pp.109.150607
- Liu, Z., Li, H., Wen, Z., Fan, X., Li, Y., Guan, R., ... Qiu, L. (2017b). Comparison of genetic diversity between Chinese and American soybean (*Glycine max* (L.) accessions revealed by high-density SNPs. *Frontiers in Plant Science*, 8. https://doi.org/10.3389/fpls. 2017.02014
- Liu, Z., Li, J., Fan, X., Htwe, N. M. P. S., Wang, S., Huang, W., ... Qiu, L. (2017a). Assessing the numbers of SNPs needed to establish molecular IDs and characterize the genetic diversity of soybean cultivars derived from *Tokachi nagaha*. *The Crop Journal*, *5*, 326–336. https://doi.org/10.1016/j.cj.2016.11.001
- Lombard, V., Baril, C. P., Dubreuil, P., Blouet, F., & Zhang, D. (2000). Genetic relationships and fingerprinting of rapeseed cultivars by AFLP: Consequences for varietal registration. *Crop Science*, 40, 1417–1725.
- Malécot, G. (1948). *Les mathématiques de l'hérédité*. Paris: Masson et Cie.
- Malek, M. A., Rafii, M. Y., Afroz, M. S. S., Nath, U. K., & Mondal, M. M. A. (2014). Morphological characterization and assessment of genetic variability, character association, and divergence in soybean mutants. *Science World Journal*, 2014, 968796. https://doi.org/10.1155/2014/968796
- Martin, J. M., Blake, T. K., & Hockett, E. A. (1991). Diversity among North American spring barley cultivars based on coefficients of parentage. *Crop Science*, *31*, 1131–1137. https://doi.org/10.2135/ cropsci1991.0011183X003100050009x
- Maton, C., Thomasset, M., Bernole, A., Remay, A., Aizac, B., & Mathis, R. (2014). *The use of molecular markers (SNP) for DUS testing*. UPOV BMT/14/10. 10p. Geneva, Switzerland: International

Union for the Protection of New Varieties of Plants. Retrieved from https://www.upov.int/edocs/mdocs/upov/en/bmt_14/bmt_ 14_10.pdf

- McDonald, M. B. (1984). Systems for crop variety identification. Retrieved from https://ir.library.msstate.edu/bitstream/handle/ 11668/14050/1984-05-McDonald.pdf?sequence=1&isAllowed=y
- Mikel, M. A., Diers, B. W., Nelson, R. L., & Smith, H. H. (2010). Genetic diversity and agronomic improvement of North American soybean germplasm. *Crop Science*, 50, 1219–1229. https://doi.org/10.2135/cropsci2009.08.0456
- Morgante, M., Brunner, S., Pea, G., Fengler, K., Zuccolo, A., & Rafalski, A. (2005). Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. *Nature Genetics*, *37*, 997–1002. https://doi.org/10.1038/ng1615
- Moss, W. W., & Hendrickson, J. A. (1973). Numerical taxonomy. Annual Review of Entomology, 18, 227–258. https://doi.org/10.1146/ annurev.en.18.010173.001303
- Nogueira, A. P. O., Sediyama, T., Cruz, C. D., Reis, M. S., Pereira, D. G., & Jangarelli, M. (2008). Novas características para diferenciação de cultivares de soja pela análise discriminante. *Ciência Rural*, *38*, 2427–2433. https://doi.org/10.1590/S0103-84782008005000025
- Noli, E., Teriaca, M. S., Sanguineti, M. C., & Conti, S. (2008). Utilization of SSR and AFLP markers for the assessment of distinctness in durum wheat. *Molecular Breeding*, 22, 301–313. https://doi.org/ 10.1007/s11032-008-9176-4
- Norris, C., Jones, H., Cockram, J., Smith, D., & Mackay, I. (2012). Final report: A potential UPOV Option 2 approach for Barley using high density genotyping. Cambridge, UK: National Institute of Agricultural Botany. Retrieved from https://cpvo.europa. eu/sites/default/files/documents/techreports/A_Potential_ UPOV_Option_Approach_for_Barley-Final_report.pdf
- Oda, M d. C., Sediyama, T., Matsuo, É., Cruz, C. D., de Barros, E. G., & da S. Ferreira, M. F. (2015). Phenotypic and molecular traits diversity in soybean launched in forty years of genetic breeding. *Agronomy Science and Biotechnology*, *1*, 1–9. https://doi.org/10.33158/ ASB.2015v1i1p1
- Ojo, D. K., Ajaya, A. O., & Oduwaye, O. A. (2012). Genetic relationships among soybean accessions based on morphological and RAPDs techniques. *Journal of Tropical Agricultural Science*, 35, 237–248.
- Palmer, R. G., Pfeiffer, T. W., Buss, G. R., & Kilen, T. C. (2004). Qualitative genetics. In H. R. Boerma & J. E. Specht (Eds.), *Soybeans: Improvement, production, and uses* (3rd. ed., pp. 137–233). Agronomy Monograph No. 16. Madison, WI: ASA, CSSA, and SSSA. https://doi.org/10.2134/agronmonogr16.3ed.c5
- Patterson, H. D., & Weatherup, S. T. C. (1984). Statistical criteria for distinctness between varieties of herbage crops. *Journal of Agricultural Science*, 102, 59–68. https://doi.org/10.1017/ S0021859600041459
- Pourabed, E., Noushabadi, M. R. J., Jamali, S. H., Alipour, N. M., Zareyan, A., & Sadeghi, L. (2015). Identification and DUS testing of rice varieties through microsatellite markers. *International Journal of Plant Genomics*, 2015, 965073. https://doi.org/10.1155/2015/ 965073
- Ramteke, R., & Murlidharan, P. (2012). Characterization of soybean (*Glycine max*) varieties as per DUS guidelines. *Indian Journal of Agricultural Sciences*, 82, 572–577.
- Ranatunga, M. A. B., Arachchi, J. D. K., Gunasekare, K., & Yakandawala, D. (2017). Floral diversity and genetic structure of tea germplasm of Sri Lanka. *International Journal of Biodiversity*, 2017, 2957297. https://doi.org/10.1155/2017/2957297

- Rasmusson, D. C., & Phillips, R. L. (1997). Plant breeding progress and genetic diversity from *de novo* variation and elevated epistasis. *Crop Science*, *37*, 303–310. https://doi.org/10.2135/cropsci1997. 0011183X003700020001x
- Rathinavel, K., Manickam, S., Sabesh, M. (Eds.). (2005). *Training* manual on DUS test in cotton with reference to PPV & FR legislation. Coimbatore, India: Central Institute for Cotton Research Regional Station. Retrieved from http://www.cicr.org.in/pdf/dus_ test_manual.pdf
- Ribeiro, C. A. G., Tanure, J. P. M., Maciel, T. E. F., & de Barros, E. G. (2013). Molecular characterization of soybean cultivars by microsatellite markers with universal tail sequence. *Pesquisa Agropecuária Brasileira*, 48, 270–279. https://doi.org/10. 1590/S0100-204X2013000300005
- Rohlf, F. J. (2008). NTSYSpc: Numerical Taxonomy System, ver. 2.21q. Setauket, NY: Exeter Publishing, Ltd.
- Rongwen, J., Akkaya, M. S., Bhagwat, A. A., Lavi, U., & Cregan, P. B. (1995). The use of microsatellite DNA markers for soybean genotype identification. *Theoretical and Applied Genetics*, 90, 43–48. https://doi.org/10.1007/BF00220994
- Sandhu, D., Ghosh, J., Johnson, C., Baumbach, J., Baumert, E., Cina, T., ... Bhattacharyya, M. K. (2017). The endogenous transposable element *Tgm9* is suitable for generating knockout mutants for functional analyses of soybean genes and genetic improvement in soybean. *PLoS ONE*, *12*, e0180732 https://doi.org/10.1371/journal. pone.0180732
- Sang-Tae, K., & Donoghue, M. J. (2008). Incongruence between cpDNA and nrITS indicates extensive hybridization within Eupersicaria (Polgonaceae). *American Journal of Botany*, 95, 1122–1135. https://doi.org/10.3732/ajb.0700008
- Santos, F. S., de Moraes Aviani, D., Hidalgo, J. A. F., Machado, R. Z., & Araújo, S. P. (2012). Evolution, importance and evaluation of cultivar protection in Brazil: The work of the SNPC. *Crop Breeding and Applied Biotechnology*, *S2*, 99–110. https://doi.org/10.1590/S1984-70332012000500011
- Scaboo, A. M., Chen, P., Sleper, D. A., & Clark, K. M. (2017). Classical breeding and genetics of soybean. In K. Bilyea, M. B. Ratnaparkhe,
 & C. Kole (Eds.), *Genetics, genomics, and breeding of soybean* (pp. 19–53). Boca Raton, FL: CRC Press Enfield.
- Shelton, A. C., & Tracy, W. F. (2017). Cultivar development in the U.S. Public Sector. Crop Science, 57, 1823–1835. https://doi.org/10.2135/ cropsci2016.11.096
- Singh, R. K., Sharma, R. K., Singh, A. K., Singh, V. P., Singh, N. K., Tiwari, S. P., & Mohapatra, T. (2004). Suitability of mapped sequence tagged microsatellite site markers for establishing distinctness, uniformity and stability in aromatic rice. *Euphytica*, 135, 135–143. https://doi.org/10.1023/B:EUPH.0000014905.10397.08
- Sneath, P. H. A., & Sokal, R. R. (1962). Numerical taxonomy. *Nature*, 193, 855–860. https://doi.org/10.1038/193855a0
- Sneller, C. H. (1994). Pedigree analysis of elite soybean lines. Crop Science, 34, 1515–1522. https://doi.org/10.2135/cropsci1994. 0011183X003400060019x
- Song, Q., Hyten, D., Jia, G., Quigley, C., Fickus, E., & Nelson, R. (2013). Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS ONE*, *8*, e54985. https://doi. org/10.1371/journal.pone.0054985
- Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., Fickus, E. W., Nelson, R. L., & Cregan, P. B. (2015). Fingerprinting soybean germplasm and its utility in genomic research. *G3: Genes, Genomes, Genetics*, 5, 1999–2006. https://doi.org/10.1534/g3.115.019000

- Song, Q., Jia, G., Quigley, C., Fickus, E., Hyten, D., Nelson, R., & Cregan, P. (2014). Soybean BARCSoySNP6K Beadchip—A tool for soybean genetics research. Poster P306. Poster session presented at. San Diego, CA: Plant and Animal Genome Conference XXII. 11–15 January. Retrieved from https://pag.confex.com/pag/xxii/ webprogram/Paper10932.html
- Song, Q. J., Quigley, C. V., Nelson, R. L., Carter, T. E., Boerma, H. R., & Strachen, J. R. (1999). A selected set of trinucleotide simple sequence repeat markers for soybean cultivar identification. *Plant Varieties and Seeds*, *12*, 207–220.
- SoyBase, (2018). SoySNP50K iSelect BeadChip. Retrieved from https://soybase.org/snps/
- Staub, J. E., Gabert, A., & Wehner, T. C. (1996). Plant variety protection: A consideration of genetic relationships. *HortScience*, *31*, 1086–1091.
- Terzić, S., Zorić, M., & Seiler, G. J. (2020). Qualitative traits in sunflower breeding: UGA-SAM1 phenotyping case study. *Crop Sci*ence, 60, 303–319. https://doi.org/10.1002/csc2.20059
- Thomasset, M., Maton, C., Bernole, A., Bazante, F., Muret, C., Aizac, B., & Remay, A. (2015). The use of SNP markers in maize DUS testing. Poster presented at the XXIIIrd EUCARPIA Maize and Sorghum Conference, 10–12 June, 2015, Montpellier, France. Retrieved from https://www.geves.fr/wp-content/uploads/Theuse-of-SNP-markers-in-maize-DUS-testing.pdf
- Thomson, R. (2013). The yield of plant variety protection. Melbourne institute of applied economic and social research, and intellectual property research institute of Australia. Melbourne, Australia: Melbourne Institute of Applied Economic and Social Research, and Intellectual Property Research Institute of Australia, The University of Melbourne Retrieved from http:// www.utas.edu.au/__data/assets/pdf_file/0020/418115/20130920_ Thomson.pdf
- Tinker, N. A., & Mather, D. E. (1993). KIN: Software for computing kinship coefficients. *Journal of Heredity*, 84, 238.
- Tommasini, L., Batley, J., Arnold, G. M., Cooke, R. J., Donini, P., Lee, D., ... Edwards, K. J. (2003). The development of multiplex simple sequence repeat (SSR) markers to complement distinctness, uniformity and stability testing of rape (*Brassica napus* L.) varieties. *Theoretical and Applied Genetics*, 106, 1091–1101. https:// doi.org/10.1007/s00122-002-1125-8
- International Union for the Protection of New Varieties of Plants. (1998a). Amendment to the draft test guidelines for soya bean. Geneva, Switzerland: International Union for the Protection of New Varieties of Plants. Retrieved from https://www.upov.int/ edocs/mdocs/upov/en/tc_34/tc_34_9.pdf
- International Union for the Protection of New Varieties of Plants. (1998b). Guidelines for the conduct of tests for distinctness, uniformity and stability: Soya Bean (Glycine max (*L.*)). Geneva, Switzerland: International Union for the Protection of New Varieties of Plants. Retrieved from https://www.upov.int/edocs/tgdocs/en/ tg080.pdf
- International Union for the Protection of New Varieties of Plants. (2002). General introduction to the examination of distinctness, uniformity and stability and the development of harmonized descriptions of new varieties of plants. Geneva, Switzerland: International Union for the Protection of New Varieties of Plants. Retrieved from https://www.upov.int/export/sites/upov/ publications/en/tg_rom/pdf/tg_1_3.pdf
- International Union for the Protection of New Varieties of Plants. (2005). Arrangements for DUS testing. Associated document to the

general introduction to the examination of distinctness, uniformity and stability and the development of harmonized descriptions of new varieties of plants. Geneva, Switzerland: International Union for the Protection of New Varieties of Plants. Retrieved from https://www.upov.int/edocs/tgpdocs/en/tgp_6_section_2

- International Union for the Protection of New Varieties of Plants. (2010). Combination of morphological distance (GAIA) with genotypic distance in the framework of "Management of the reference collection." Geneva, Switzerland: International Union for the Protection of New Varieties of Plants. Retrieved from https://www. upov.int/edocs/upov/en/twc/28/twc_28_28.pdf
- International Union for the Protection of New Varieties of Plants. (2011). Possible use of molecular markers in the examination of distinctness, uniformity, and stability (DUS). Geneva, Switzerland: International Union for the Protection of New Varieties of Plants. Retrieved from https://www.upov.int/edocs/infdocs/en/ upov_inf_18_1.pdf
- International Union for the Protection of New Varieties of Plants. (2015). Examining distinctness. Associated document to the general introduction to the examination of distinctness, uniformity, and stability and the development of harmonized descriptions of new varieties of plants. Geneva, Switzerland: International Union for the Protection of New Varieties of Plants. Retrieved from https://www. upov.int/edocs/tgpdocs/en/tgp_8.pdf
- International Union for the Protection of New Varieties of Plants. (2016a). *Report by the Community Plant Variety Office (CPVO) to the UPOV BMT meeting in Moscow, 24–27 May 2016 on CPVO IMOD-DUS*. Geneva, Switzerland: International Union for the Protection of New Varieties of Plants. Retrieved from http://www.upov.int/ edocs/mdocs/upov/en/bmt_15/bmt_15_27.pdf
- International Union for the Protection of New Varieties of Plants. (2016b). *Reports on developments in UPOV concerning biochemical, and molecular techniques*. Geneva, Switzerland: International Union for the Protection of New Varieties of Plants. Retrieved from https://www.upov.int/edocs/mdocs/upov/en/bmt_15/bmt_ 15_2.pdf
- International Union for the Protection of New Varieties of Plants. (2019). Data Processing for the Assessment of Distinctness and for Producing Variety Descriptions. TWP/3/10. Geneva, Switzerland: International Union for the Protection of New Varieties of Plants. Retrieved from https://www.upov.int/edocs/mdocs/upov/ en/two_51/twp_3_10.pdf
- Urrea, K., Rupe, J., Chen, P., & Rothrock, C. S. (2017). Characterization of seed rot resistance to *Pythium aphanidermatum* in soybean. *Crop Science*, 57, 1394–1403. https://doi.org/10.2135/cropsci2016. 08.0669
- USDA-Agricultural Marketing Service, (2019). Plant Variety Protection Office—Scanned and redacted issued certificates. Washington, DC: USDA-Agricultural Marketing Service. Retrieved from https://apps.ams.usda.gov/CMS/
- Van Beuningen, L. T., & Busch, R. H. (1997). Genetic diversity among North American spring wheat cultivars: I. Analysis of the coefficient of parentage matrix. *Crop Science*, 37, 570–579. https://doi. org/10.2135/cropsci1997.0011183X003700020043x
- Van Ettekoven, K. (2017). Room for the use of bio molecular techniques in the DUS process? *European Seed Newsletter*,
 4. Retrieved from https://european-seed.com/2017/09/room-use-bio-molecular-techniques-dus-process/
- Varala, K., Swaminathan, K., Li, Y., & Hudson, M. E. (2011). Rapid genotyping of soybean cultivars using high throughput sequenc-

ing. *PLoS ONE*, *6*, e24811. https://doi.org/10.1371/journal.pone. 0024811

- Vieira, E. S. N., Pinho, E. V. R.V., Carvalho, M. G. G., & Silva, P. A. (2009). Caracterização de cultivares de soja por descritores morfológicos emarcadores bioquímicos de proteínas e isoenzimas. *Revista Brasileira de Sementes*, *31*, 86–94. https://doi.org/10.1590/S0101-31222009000100010
- Wagner, C. K., & McDonald, M. B. (1981). Identification of soybean (Glycine max (L.) Merrill) cultivars using rapid laboratory techniques. Res. Bull. 1133. Wooster, OH: Ohio Agricultural Research and Development Center. Retrieved from https://kb.osu.edu/bitstream/handle/1811/62955/OARDC_ research_bulletin_n1133.pdf?sequence=1&isAllowed=y
- Wallace, M. (2017). DUS testing and molecular techniques: A view from the UK plant breeder's rights office. *European Seed Newsletter*, 4. Retrieved from https://european-seed.com/2017/09/dustesting-molecular-techniques-view-uk-plant-breeders-rightsoffice/
- Wang, L. X., Li, H. B., Gu, T. C., Liu, L. H., Pang, B. S., Qiu, J., & Zhao, C. P. (2014). Assessment of wheat variety stability using SSR markers. *Euphytica*, 195, 435–452. https://doi.org/10.1007/s10681-013-1006-z
- Wang, S., & Lu, Z. (2006). Genetic diversity among parental lines of Indica hybrid rice (Oryza sativa L.) in China based on coefficient of parentage. *Plant Breeding*, *125*, 606–612. https://doi.org/10.1111/ j.1439-0523.2006.01268.x
- Wilcox, J. R., Athow, K. L., LLaviolette, F. A., Abney, T. S., & Richards, T. L. (1979). Registration of Wells II soybean. *Crop Science*, 19, 296. https://doi.org/10.2135/cropsci1979.0011183X001900020040x
- Wurtenberger, G. (2017). The object of plant variety protection and the scope deriving therefrom. Retrieved from https://cpvo.europa.eu/en/news-and-events/articles/objectplant-variety-protection-and-scope-deriving-therefrom
- Yan, F., Di, S., Rodas, F. R., Torrico, T. R., Murai, Y., Iwashina, T., ... Takahashi, R. (2014). Allelic variation of soybean flower color gene W4 encoding dihydroflavonol 4-reductase 2. BMC Plant Biology, 14, 58. https://doi.org/10.1186/1471-2229-14-58
- Yates, J. L., Boerma, H. R., & Fasoula, V. A. (2012). SSR-marker analysis of the intracultivar phenotypic variation discovered within 3 soybean cultivars. *Journal of Heredity*, 103, 570–578. https://doi. org/10.1093/jhered/ess015
- Yoon, M. S., Song, Q. J., Choi, I. Y., Specht, J. E., Hyten, D. L., & Cregan, P. B. (2007). BARCSoySNP23: A panel of 23 selected SNPs for soybean cultivar identification. *Theoretical and Applied Genetics*, 114, 885–899. https://doi.org/10.1007/s00122-006-0487-8

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Achard F, Butruille M, Madjarac S., et al. Single nucleotide polymorphisms facilitate distinctness-uniformity-stability testing of soybean cultivars for plant variety protection. *Crop Science*. 2020;1–24.

https://doi.org/10.1002/csc2.20201