



BMT/15/14

ORIGINAL: English

DATE: May 17, 2016

**INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS**

Geneva

**WORKING GROUP ON BIOCHEMICAL AND MOLECULAR TECHNIQUES  
AND DNA-PROFILING IN PARTICULAR****Fifteenth Session****Moscow, Russian Federation, May 24 to 27, 2016****COMPARISON OF GENOTYPIC AND EXPRESSION DATA TO DETERMINE DISTINCTNESS AMONG  
INBRED LINES OF MAIZE FOR GRANTING PLANT BREEDERS' RIGHTS***Document prepared by experts from Seed Association of the Americas**Disclaimer: this document does not represent UPOV policies or guidance***INTRODUCTION**

1. In UPOV, the examination of distinctness, uniformity and stability (DUS) is based on morphological and physiological characteristics. With the rapid evolution and increase in cost effective use of molecular markers, these data are increasingly playing a role in identifying varieties both post their grant of protection, and in the granting of protection as a method of examining morphological and physiological DUS characteristics that satisfy the UPOV criteria for characteristics. Further, molecular markers are also being used to help manage variety collections that have become unmanageable due to sheer size which is taxing on resources (UPOV, 2011). The interest in molecular markers goes beyond the fact that they exist, but rather they offer real advantages in DUS testing and the desire to harmonize systems.
2. The Working Group on Biochemical and Molecular Techniques and DNA-Profiling in Particular (BMT) was specifically instituted to "study DNA profiling in connection with plant breeders' rights and to coordinate the development and harmonization of DNA analysis in the UPOV member States" (UPOV, 1993a). The use of DNA per se continues to be scrutinized and debated: "Some experts were of the view that proof of the presence of a certain DNA in itself was not enough and that it was important that there be an expression of that certain DNA." .... "Others questioned why the presence of a mere morphological feature of a plant should be thought to be more useful for description/identification purposes than the presence of an apparently non-translated segment of DNA." (UPOV, 1993b). Interestingly, in this regard the definition of "variety" in the 1991 Act of the UPOV Convention refers to the "...expression of characteristics resulting from a given genotype." (UPOV, 1991)
3. With the goal to help resolve these positions on use of characteristics and indeed to improve the quality of variety protection, we used 10 publicly available inbred lines of maize to evaluate and compare practically the relative potential of Single Nucleotide Polymorphism (SNP) profiles with data generated using each of three expression based characterization methods: 1. RNA transcription profiles, 2. metabolome product profiles, and 3. morphological traits. We also compared results in the context of the stated pedigrees of each inbred line. We used both the range of expression and robustness of each dataset to develop an index to allow for a comparison of the practical utility of each methodology for measuring distances among these maize lines.

## MATERIALS

4. Ten publicly available inbred lines developed in the United States of America were used in the study representing 3 important heterotic groups that are widely used in temperate maize (Table 1).

Table 1.

Variety Name	PI Accession Number	PVP Number	PVP Applicant Name	Pedigree in PVP
Mo17	PI 558532	N/A	Public	C1187-2 x C103
Seagull Seventeen	PI 600751	7900077	Rothermel Seed Company	Mo17 x Unknown
LH51	PI 600955	8200062	Holden's Foundation Seeds, Inc.	Mo17
740	PI 601489	8800028	Novartis Seeds, Inc.	Mo17 x Mexican Deep Kernel
207	PI 601005	8300144	Pioneer Hi-Bred International, Inc.	G3BD2 x G3RZ1
Q381	PI 601190	8500098	Quality Research Associates	Off-type from Pioneer hybrid 3369
IBB15	PI 601458	8700196	DeKalb-Pfizer Genetics	J6 x W70884
L 135	PI 601727	8900202	Lifaco Seed Corporation	P3901 x W117
B73	PI 550473	N/A	Public	Iowa SSS C5
DJ7	PI 601191	8500086	Edward J. Funk & Sons, Inc.	B73 x BS16

## METHODS

### SNP Profiling

5. Lyophilized leaf tissue from 4 individual plants from each maize line was bulked for DNA extraction. The assay utilized the Maize SNP50 BeadChip (Illumina®) (Illumina, Inc., San Diego, CA, USA) to generate SNP profiles. A quality control step was done where SNPs with <10% heterozygotes and >80% data present across the 10 inbreds were retained for analysis. In all, 52,406 SNP markers were used in the study.

### RNA Transcription Profiling

6. Two experiments were conducted in October, 2011 and February, 2012. Plants were grown in a greenhouse and sampled at the 2 collared leaf stage. Samples were collected at two time points; 8 a.m. and again at 2 p.m. Sequences were generated on the Illumina HiSeq 2500 (Illumina®) (Illumina, Inc., San Diego, CA, USA) system with appropriate reagents and protocols. Data quality was assessed by sample correlation and hierarchical clustering of replicates.

### Metabolome Profiling

7. Nine individual plants were sampled for each respective inbred from 2 and 4 week old greenhouse plants. Gas Chromatography Time of Flight – Mass Spectrometry was used to measure metabolites as described by (Asiago et al., 2012).

### Morphological Traits

8. Each inbred was characterized using the morphological traits required by the United States Plant Variety Protection Office (US PVP Office) for DUS examination. Traits were collected from 4 different environments for all 10 inbreds.

## EVALUATION INDICES

9. Range of Expression (variability) and robustness across environments (precision) were used as the test metrics to compare characterization methods based upon previous research by Law et al. Range of expression is (maximum observed value of similarity) – (minus minimum observed value of similarity). Robustness across environments was estimated by measuring the % of inbred lines for which 75% or more

sample replicates clustered together in a hierarchical clade. The products of the two components are expressed as a discrimination index % to compare the different characterization methodologies, where a value of 0 is useless and a value of 100 is ideal.

## RESULTS AND DISCUSSION

10. The range of expression (variability), robustness across environments (precision), and discrimination index are presented in Table 2.

Table 2.

Characterization Method	Variability (%)	Precision (%)	Discrimination Index (%)
SNPs	51	70	35.7
RNA Transcription	26	70	18.2
Metabolomics	51	0	0
Morphology	98	20	19.6

11. The central dogma of genetics represents the primary biological components and processes of the cell leading from DNA to RNA and from RNA to protein. SNP data represent the DNA or genotypic, RNA transcription represents the expression of genes within DNA, and metabolomics represents the many diverse proteins produced from the RNA. Ultimately a plant's morphology is derived from the metabolomics condition.

12. Gene expression does not occur in a vacuum, but rather responds and interacts with the environment the organism lives in. With this in mind, of the four characterization methods, SNPs had the greatest discrimination index nearly twice that of the next closest method. This high value is contributed by the combination of very high precision (Figure 1) and moderate variability. Very high precision is due to the lack of environmental interactions, including very low rates of laboratory error, since the DNA sequence itself is interrogated. RNA transcription performed poorest in terms of variability and shared with SNPs the highest value for precision (Figure 2) resulting in, at best a moderate discrimination index. Metabolomics had a good range of variability, but proved to be completely unreliable for precision (Figure 3), and thus deemed useless to meaningfully and repeatedly discriminate varieties. Lastly, morphology precision (Figure 4) performed poorly leading to only a moderate discrimination value, however variability was very high which is driven in part by the effects of interactions with the environment, which would then cause this value of discrimination to be an over estimate.

Figure 1. Associations among inbred lines on the basis of comparing SNP profiles using multivariate analysis of pair-wise distance data from 52,406 SNPs.

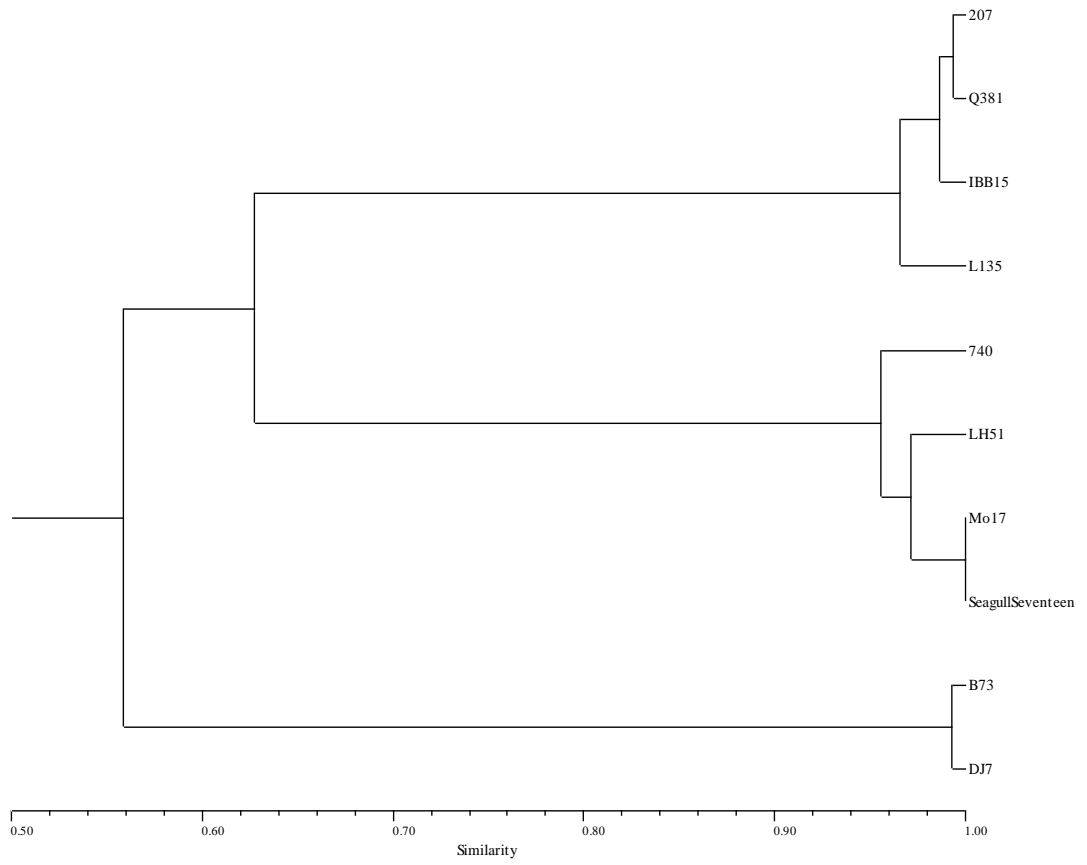


Figure 2. Associations of 4 replicate samples and 2 harvest times that preferentially clustered following multivariate analysis of 24,439 transcripts.

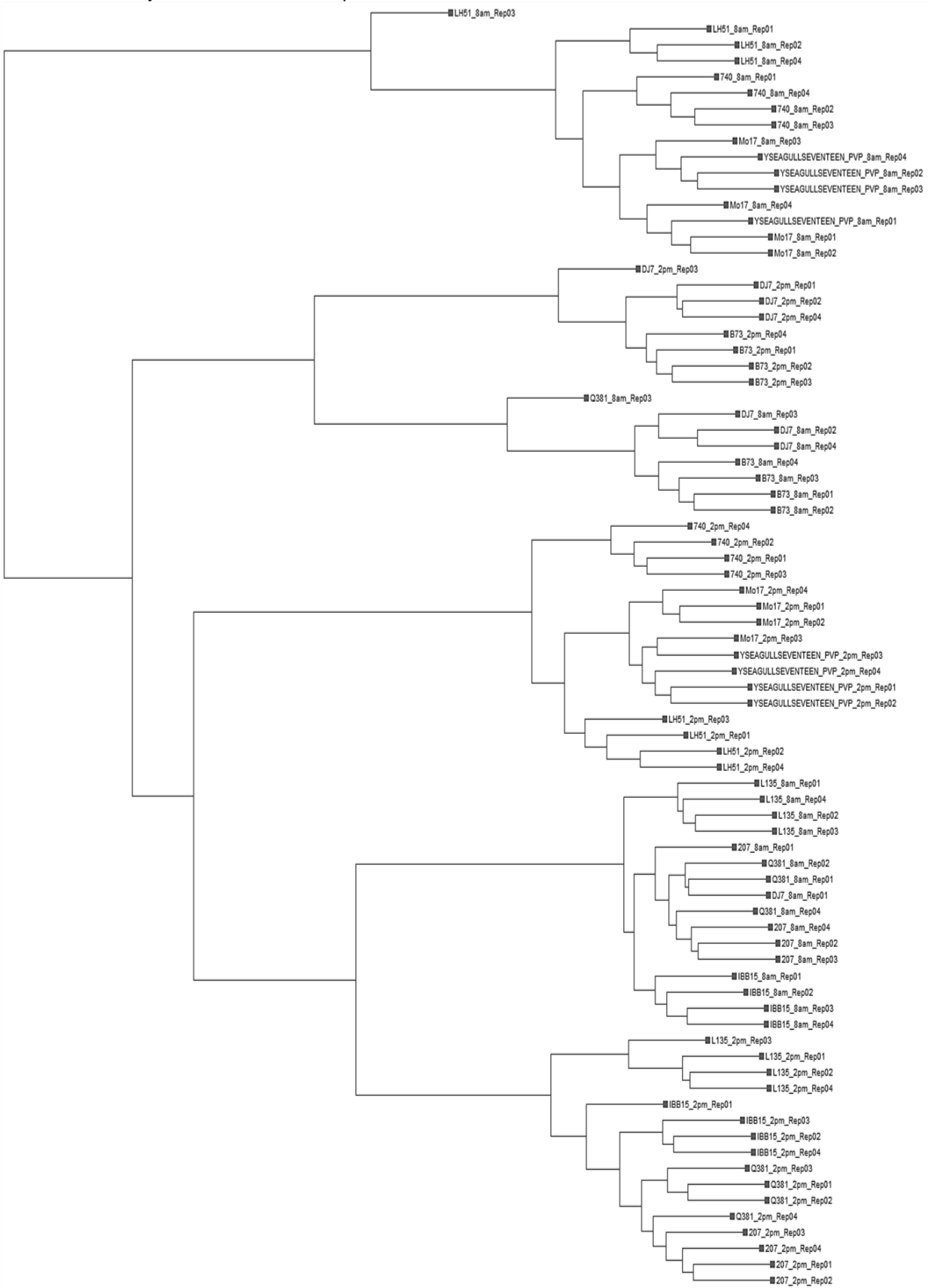


Figure 3. Associations among samples on the basis of multivariate analysis of 9,424 metabolite features using metabolomics data from experiments on 2 week and 4 week old plants.

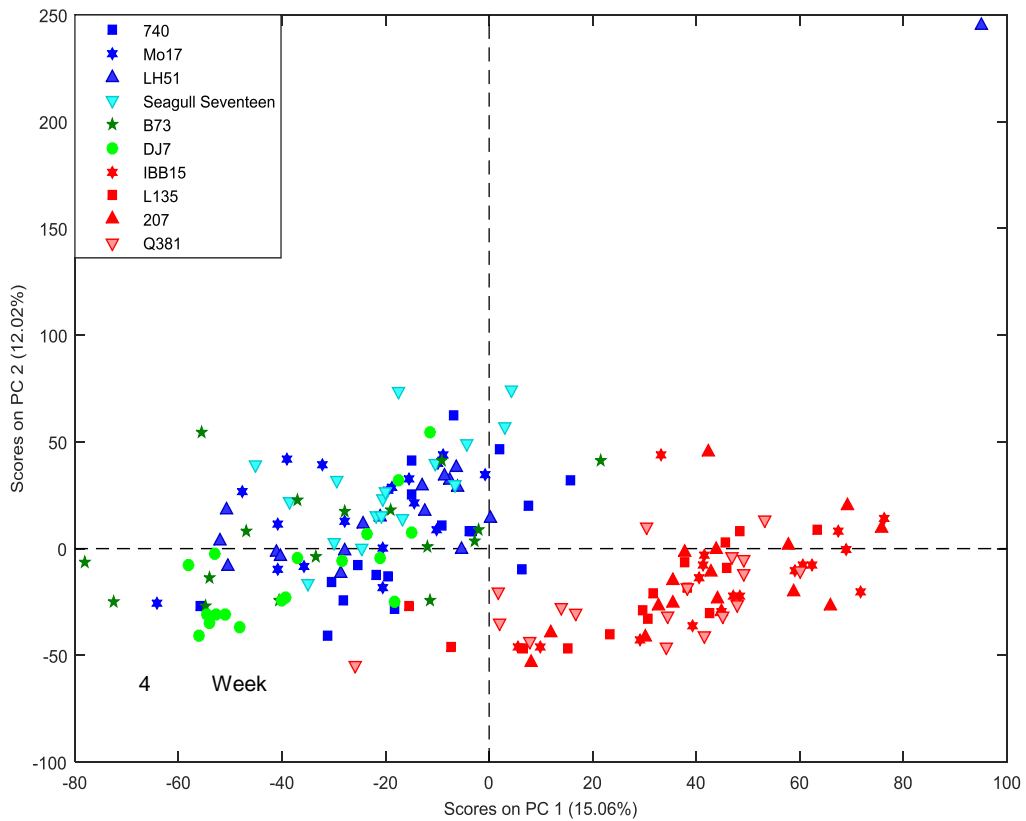
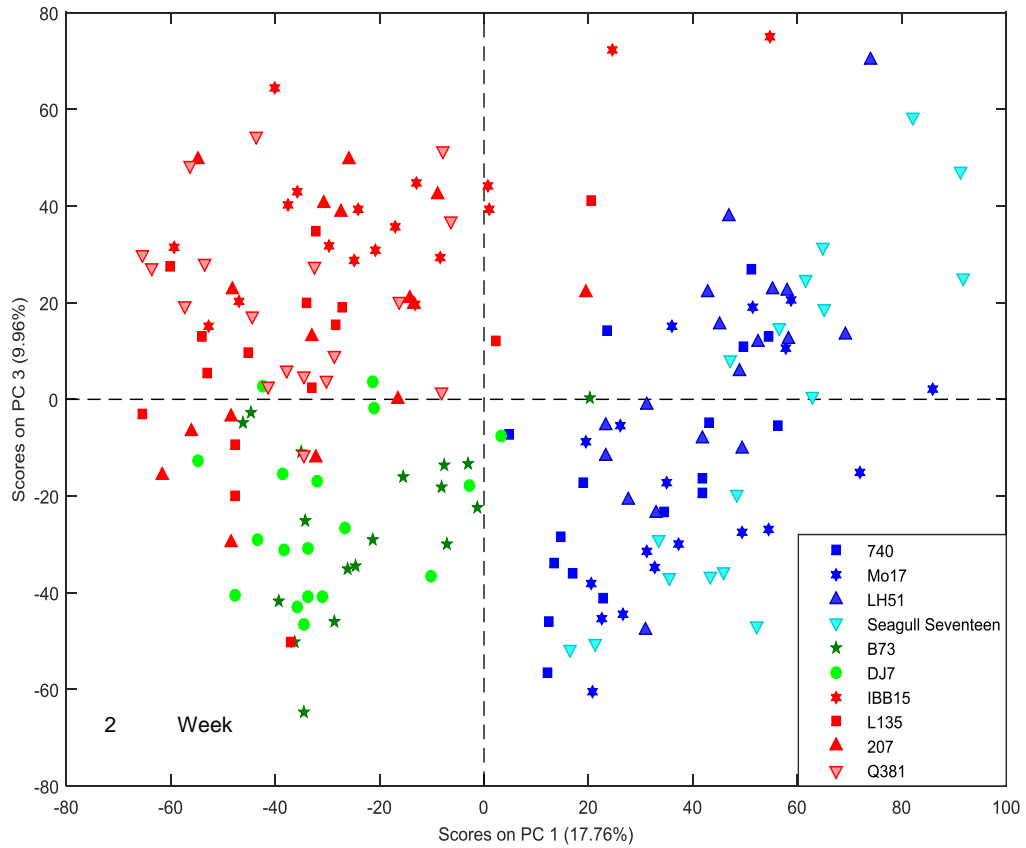
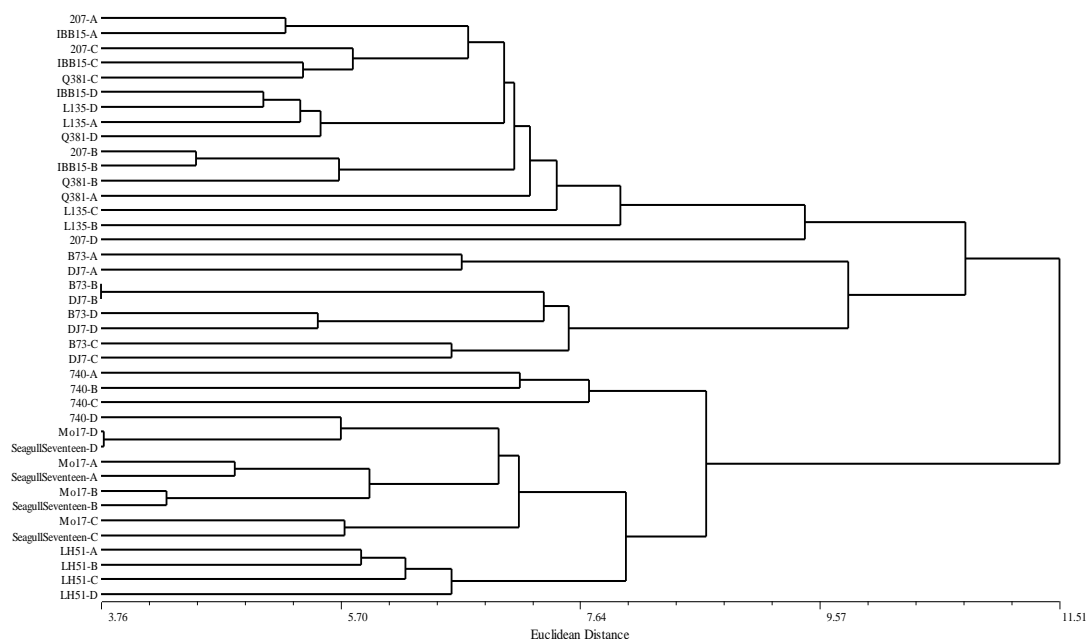


Figure 4. Associations among inbred lines according to multivariate analysis of morphological data obtained in each of 4 environments (A, B, C, and D).



13. SNPs and RNA transcription had the same degree of precision, but RNA transcription data had only 50% of the level of discrimination shown by SNPs. However, in terms of both precision and in showing associations among inbred lines, SNP data showed the same results as expressed RNA transcription data. In practical terms therefore, SNP data are equally reliable as expressed RNA data yet more discriminative due to the range of variability SNPs are able to exhibit. Furthermore, SNP data are far more cost effective and methods can be much more easily standardized than for RNA transcription data.

14. The definition of a variety in the 1991 Act of the UPOV Convention refers to “expression of characteristics resulting from a given genotype” (UPOV, 1991). These data demonstrate that 1. Associations among inbred lines on the basis of SNP and RNA expression data are the same, at least for inbred lines that are 97.2% similar by SNPs, and 2. SNP genotyping is superior to RNA, protein, and morphological expression-based characterization methods to discriminate inbreds, due to high precision (shared with RNA expression data) coupled with an ability to reveal a high degree of variability due to freedom from the hindrances of Genotype X Environment interactions and Environment effects. No other methodology we evaluated exhibited high scores for both the degree to reveal variability among inbred lines and the degree of precision.

## REFERENCES

Asiago, V.M., J. Hazebroek, T. Harp, and C. Zhong. 2012. Effects of Genetics and Environment on the Metabolome of Commercial Maize Hybrids: A Multisite Study. *J. Agric. Food Chem.*, 2012, 60:11498–11508.

Law, J.R., S.R. Anderson, E.S. Jones, B. Nelson, E. Mulaosmanovic, B. Hall, and J. S.C. Smith. 2011. Approaches to improve the determination of eligibility for plant variety protection: I. Evaluation of morphological characteristics. *Maydica* 56-1713.

UPOV, 1993a. Report adopted by the technical committee from the 28th session, October 21-23,1992 in Geneva. TC/28/6. UPOV, Geneva, Switzerland.

UPOV, 1993b. Report adopted by the Working Group on Biochemical and Molecular Techniques and DNA-profiling in particular. BMT/1/4. UPOV, Geneva, Switzerland.

UPOV. 1991. International convention for the protection of new varieties of plants of December 2, 1961, as revised at Geneva on November 10, 1972, on October 23, 1978, and on March 19, 1991. UPOV, Geneva.

UPOV. 2011. Possible use of molecular markers in the examination of Distinctness, Uniformity, and Stability (DUS). UPOV/INF/18//1 with 4 annexes. UPOV, Geneva, Switzerland. p. 26.

Barry Nelson<sup>1</sup>, Stephen Smith<sup>2</sup> et al.

<sup>1</sup>DuPont Pioneer, Johnston, IA. USA

<sup>2</sup>Seed Science Center, Iowa State University, Ames, IA. USA

[End of document]