Original Paper Open Access

Molecular analysis of genetic diversity in a Texas maize (*Zea mays* L) breeding program

Schuyler D Smith[1], Seth C Murray[1]*, Elliot Heffner[2]

[1]Department of Soil and Crop Sciences, Texas A&M University, College Station, TX 77845, USA
[2]DuPont Pioneer, Johnston, Iowa Johnston, IA 50131, USA
*Corresponding author: E-mail: sethmurray@tamu.edu

Abstract

The Texas maize (*Zea mays* L) breeding program at Texas A&M University has been unique among breeding programs for the incorporation of diverse germplasm from a wide range of origins into elite inbred lines. The Texas program, situated in a subtropical environment, has found beneficial traits in maize of tropical origin beyond what is available in the temperate material commonly used in the far more productive Midwestern region of the United States. To date, no molecular studies had been conducted to make any quantitative differentiations between the genetic diversity in the germplasm developed in the Texas program or comparisons to the germplasm available from the Midwest. In this study, a molecular characterization of genetic diversity was performed. A unique set of 266 elite Texas lines were genotyped using 766 single nucleotide polymorphism markers, this was then combined with data published in a previous study focusing on ex-PVP lines released by private companies. The two data sets combined had 380 genotypes with 635 markers. It was determined that there were five subpopulations of material in this combined set as demonstrated by population structure. The data suggested that the array markers, designed to cluster the Midwestern heterotic groups, did not discriminate this exotic material well and/or that the Texas heterotic pools were not well supported. We conclude that the majority of Texas program material is a novel population, genetically dissimilar to Midwest temperate material, and would be a useful source of unique genetics for other maize breeding programs.

Keywords: molecular characterization, subtropical maize, heterotic groups, genetic diversity, STRUCTURE

Introduction

The Texas maize breeding program (Rogers and Collier, 1952) has been focused on improving maize in a subtropical environment using germplasm unique to that in temperate programs in the US. The germplasm used has been acquired from programs throughout the Americas, in particular the tropical and subtropical regions of Central and South America such as CYMMIT, Bolivia, Argentina, Colombia, and Peru. Though this germplasm has been incorporated and used with strong results in the program's breeding (Mayfield et al, 2012; Betran et al, 2004; Barerro-Farfan et al, 2015), little molecular information has been gathered on the genetic diversity and relationships of these pedigrees. It is not possible to investigate diversity of hybrids currently used in production, but it can be investigated via expired plant variety protected (ex-PVP) lines which are derivatives of current industry lines (Mikel and Dudley 2006; Mikel, 2008). However, many Texas elite lines do possess any pedigree relationships with material that is used elsewhere in the US maize industry.

Since the lines have such diversity it would serve better to do a molecular study as opposed to a pedigree analysis. It has been found that pedigree information provides estimates for genetic composition with significantly less accuracy than genotyping data. This is often a result of missing, incomplete, or inaccurate pedigree information (Munoz et al, 2014); as is the case in the Texas maize breeding program. Both type I and type II error are reduced when using marker information compared to pedigree records (Yu et al, 2005).

Why the Texas maize breeding program material is useful beyond Texas

Recent goals being put forward are to double yields by the year 2050, to meet the demands of the growing world population (Alexandratos and Bruinsma, 2012). Although this ambitious goal is already behind schedule (Ray et al, 2013), it will continue to face new obstacles from a changing climate never before seen by breeders in the major areas of production (Chapman et al, 2012; Ceccarelli et al, 2010). One of the most important known changes will be temperature, with average world temperature expected to be 4 degrees higher by 2050 (Hayhoe et al, 2010, Chapman et al, 2012). This could mean that areas such as Iowa and the US Corn Belt could be facing summers similar to what Texas experiences now. Heat stress is already a major factor in yield potential of corn from season to season, and the Corn Belt may benefit from subtropical and tropical germplasm as a source of heat tolerance.

Along with the increasing global temperatures,

 RECEIVED 07/17/2015

areas of historically temperate climates will be presented with the emergence of biological threats that have never before been an issue. One of the most important threats that increases with heat, drought and other stresses is aflatoxin, a carcinogenic mycotoxin produced in the kernel by the fungus *Aspergillus flavus*. Decreased susceptibility to Aflatoxin accumulation is a major focus of the Texas maize breeding program, in addition to yield, heat stress tolerance, and drought stress tolerance (Barerro et al, 2015; Mayfield et al, 2011). In the case of these new diseases and pests it would be highly beneficial to have access to lines and information on genetic similarity that would assist in quick assimilation of the most adapted material produced to deal with such adversities.

Tropical, also called exotic, germplasm has a wealth of diversity that is not homologous with temperate germplasm since the wild progenitor species of maize is from the tropics (Reif et al, 2004; Tarter et al, 2004; Buckler et al, 2006). A large aspect of the appeal to the diversity that tropical material brings is its multitudes of resistance to diseases (De Leon and Pandey, 1989; Khairallah et al, 1998; Poland et al, 2011). However, it has been suggested that the two major groups in maize world-wide are divided as temperate and tropical (Yan et al, 2009; Lu et al, 2009), and each set of material is not easily grown in the other's environment. The difficulty is that with the non-homologous genomes there are risks of higher linkage drag of undesirable traits and difficulty in combining target regions.

Very few breeding programs in the US work with tropical material. The most well published and successful program has been that initially led by Dr M Goodman out of North Carolina State University. Their studies found that tropical material can be crossed with temperate material with relative success and without any significant detriment to yield (Tallury and Goodman, 1999). This motivation led to the successful and ongoing USDA Genetic Enhancement of Maize (GEM) program which seeks to increase the diversity of industry temperate maize through the addition of tropical diversity (Pollak and Salhuana, 2001; Pollak, 2003). This approach has also been used in the Texas maize breeding program, however the program has also successfully pursued pedigree selection from tropical x tropical crosses and directly from tropical populations.

Although heterotic groups are carefully maintained in Midwestern corn breeding and consist of BSSS, NSS, and Iodents (Melchinger et al, 1991; Nelson et al, 2008) it is not clear where the lines selected in Texas from tropical x temperate and tropical x tropical crosses would fit in relation to any of those accepted groups. Some information has been gained from using industry tester lines but only two testers were initially used, the stiff stalk LH195 (Holdens, 1991) and the non-stiff stalk LH287 (Holdens, 2002). Subsequently, many of the best Texas lines have been crossed with additional commercial testers and demonstrated results that confirmed initial results from these two testers.

With keen understanding of the vulnerability of maize, Dr Major Goodman and colleagues sought to use molecular markers to characterize both commercial germplasm and their own. In 2008, Nelson et al compared temperate public lines from Dr Goodman's program at North Carolina State University to temperate commercial ex-PVP lines with the objective to demonstrate genetic relatedness and groupings. Breeders in sub-tropical regions and even the Midwestern states would likewise benefit from better understanding of Texas lines for predicting heterotic group membership and successful crosses. The objectives of this study were to 1) characterize the genetic diversity in the Texas corn breeding program, 2) characterize the lines of the Texas corn breeding program in comparison to ex-PVP and public lines from the Midwest, 3) gain insight into heterotic groupings in maize, and 4) show any pedigree misclassifications of lines that may have occurred over the years of line development.

Materials and Methods

Germplasm

The novel data set used in this study was comprised of 263 breeding lines (set 1) from the Texas corn breeding program that were selected along with 13 ex-PVP Midwest inbreds (including LH82, LH195, B104, B73), totaling 276 entries. Domestic lines originating from the Midwest (LH52(Mo17), B73, B104; Hallauer et al, 1997), North Carolina (NC300; Goodman et al, 1991), and Texas (Tx714; Betran et al, 2004), were chosen for importance and because they were in the pedigrees of some newly developed Texas lines. Complete genotyping data was collected on 266 lines which were used for all analyses from this set. Set 1 was combined (Supplementary Table 1) with 114 lines (set 2) selected and published by Nelson et al (2008). Nelson et al (2008) used 17 public inbreds and 92 ex-PVP lines, with five representatives of a B73/Mo17 hybrid.

Table 1 - Calculated variation explained by the first 12 eigenvectors using the cmd function in R.

Eigenvalue	Total explained	Additional
1	11%	11%
2	18%	7%
3	22%	4%
4	25%	3%
5	28%	2%
6	30%	2%
7	32%	2%
8	34%	2%
9	36%	2%
10	38%	2%
11	39%	2%
12	41%	1%

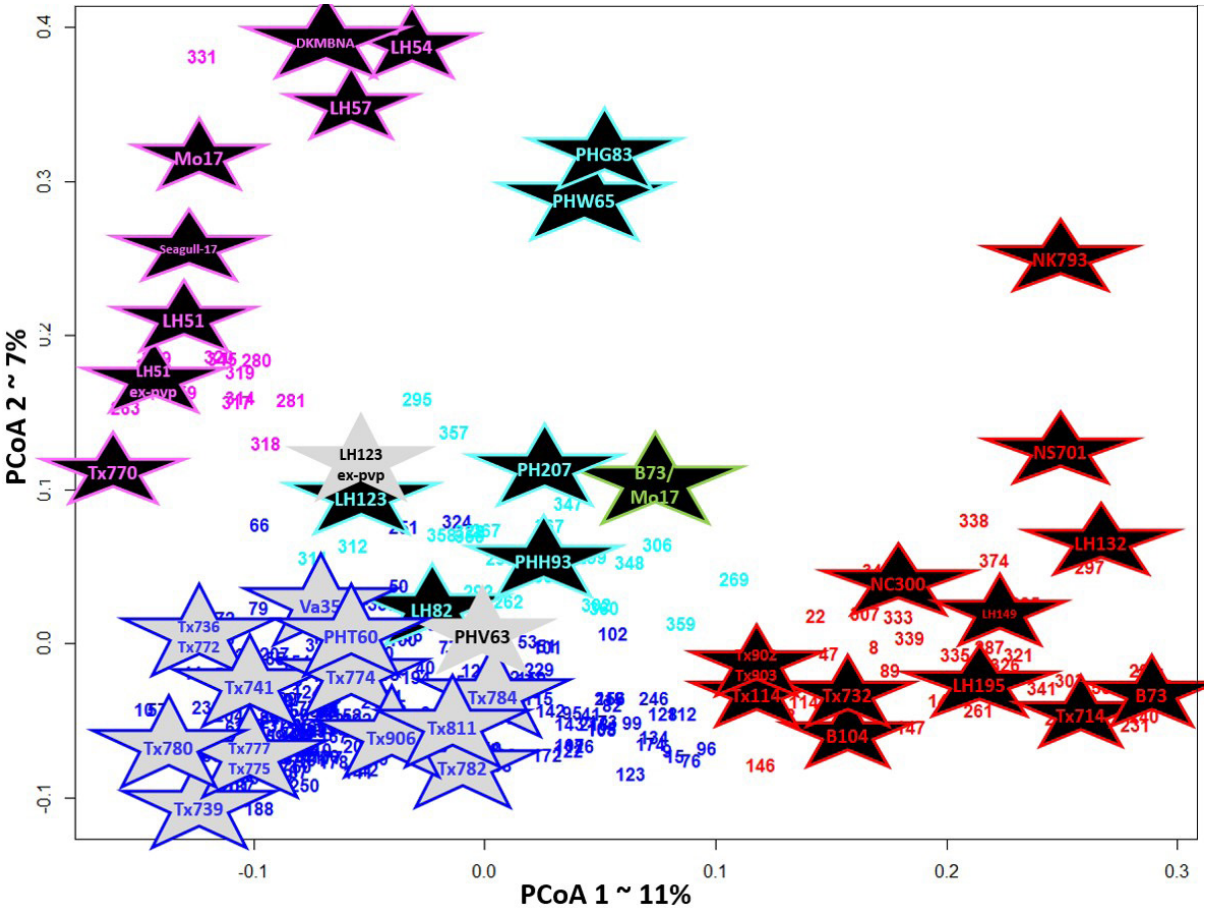

Figure 1 - Graphical representation showing the first by second eigenvector of the PCoA analysis explaining 18% of the variation in the data. Individuals are represented by their entry number in the dataset (Supplementary Table 1), with some individuals of higher interest labeled with pedigree names. Four clusters and the B73/Mo17 hybrid were identified; on the higher (right) end of the x-axis lies B73 and its derivatives. On the left end of the x-axis are the NSS lines. Mo17, a Non-Stiff Stalk is present on the higher end of the y-axis. PH207, an Iodent, is in the center of the figure. Colors denote the five grouping by Structure.

Genotyping

Seed for each of the 276 lines selected for the study were potted individually and grown for 15 days after germination. Leaf tissue was then individually sampled and placed into 96-well plates. DNA extraction and genotyping was performed by DuPont Pioneer (Johnston, IA) using the Illumina GoldenGate assay (Fan et al, 2004) with the same 768 public SNP markers used by Nelson et al (2008). These markers were selected for their proven universal heterozygosity values > 0.2 in most groups of maize lines (Nelson et al, 2008). Because of this attribute, it stands to reason that they were a good set to use to screen Texas lines, having a very different background than the vast majority of Midwest lines, like the ones Nelson et al (2008) screened. In total, 766 SNP markers (Supplementary Table 3) were successfully scored and provided by DuPont Pioneer for the set 1 Texas lines (Supplementary Table 1).

Data analysis 1

The software program PowerMarker (Liu and Muse, 2005) was used to compute allelic frequencies and calculate genetic distances for each unique pair using the algorithms of Nei 1973, which was chosen based on its conservative estimation of distance (Nei, 1987). These values for genetic distance were used to create a distance matrix, which was then input into the statistical analysis software program R for principal coordinate analysis (PCoA). Using the genetic distance has advantages over the raw marker calls in this data set since there are more markers than individuals and some missing data (Rohlf, 1972). The R function cmd was used to calculate eigenvalues for the PCoA. R was then used to plot the first by second principal coordinates giving a graphical representation of the relations between all of the 380 maize lines. The software STRUCTURE v 2.3.4 (Pritchard et al, 2000) was used for cluster analysis (Everrit, 1980) of the data. The majority of options were left to their default settings as advised by Pritchard et al (2000). The parameters were run on non-hierarchical analysis using an admixture model (Balding and Nichols, 1995) with K = 1 - 8, with burnin and MCMC values set at 10,000 for 20 iterations, which has been shown to be

sufficient (Evanno et al, 2005). With the preliminary results, the optimal K value was calculated with posterior probability analysis as demonstrated in Evanno et al (2005). The cluster analysis data was attributed to the principal coordinate analysis in R and plotted on the same principal coordinate plot. Influential lines were identified on the graphs to identify cluster families.

Results and Discussion

A total of 741 markers were used for the set 1 as 25 of the markers returned no data across any lines. Eight of the 276 Texas lines (Supplementary Table 1) were omitted for heterozygosity greater than 8%, and two returned no data from the genotyping, which left 266 lines from set 1 for analysis. When the genotyping data from the set in this study was combined with that data from Nelson et al (2008), there were 635 shared markers (Supplementary Table 1) across 380 lines. On average, each of the 635 markers used in the analysis returned data for 365 of the 380 lines (96%). Each of the 380 lines returned allele data for 606 markers, on average, a 95% return (Supplementary Table 1). The average minor allele frequency was 26%, and well distributed between 1% and 50% (Supplementary Figure 1).

The distance matrix (Supplementary Table 2) developed from the 380 lines was validated by comparing lines of known pedigrees and was determined to have results consistent with our expectations for all closely related lines (low distance values). The PCoA explained part of the variation present in the maize genotype set but was lower overall than expected

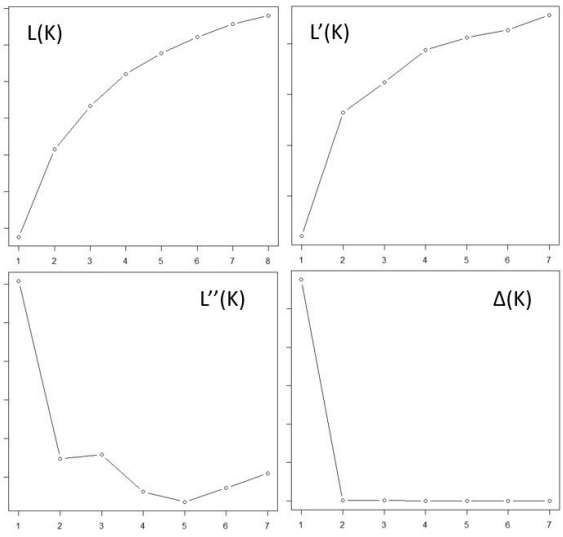

Figure 2 - posterior probability analysis of the log likelihood output from STRUCTURE plotted using the statistical software program R. L(K) is the average of the log likelihood values given from structure with half their variance subtracted out. L'(K) is the mean distance between each level of K = L(K)-L(K-1). L"(K) = L'(K+1)-L'(K), the difference between each level of L'(K). ΔK = L"(K)/s(L(K)).

(Table 1). Different combinations of principal coordinates one to five were evaluated based on our knowledge of the pedigrees but no combination was found to offer any better explanation or show any additional useful patterns of the dataset greater than that of coordinate one versus two.

The graphical representation of the PCoA (Figure 1) revealed the dispersion of the entries based from the PCoA. From the individual entries it was determined that the first principal coordinate had separated the Iowa Stiff Stalk Synthetics (BSSS) and Non Stiff Stalks (NSS), with the vast majority of the entries being in the latter. The second principal coordinate gave no clear explanation of what specific quality was being differentiated; however, it was evident that it separated the lines from the Texas breeding program apart from the Midwest NSS lines and showed a larger differentiation between the tropical lines and Mo17 than the tropical lines and B73.

Based on the genotype data from the marker set, it is likely that the second principal component differentiated the NSS lines from everything else. Another observed trend was that the Iodent heterotic group, represented by PH207 and PHH93, clustered between the BSSS and NSS clusters but above the Texas cluster. This was not surprising because the Iodents were derived from the Reid Yellow Dent open pollinated variety which was also the origin of the 75% of the BSSS parents that created the stiff stalks and most of the parents of the «Lancaster» non-stiff stalks (Lee and Tracy, 2012).

Because PCoA is not model based and does not provide evidence on group membership, STRUCTURE was used to estimate the optimal number of populations that could be described by the data. The posterior probability analysis from STRUCUTRE showed that the best value of K, or true number of clusters represented in the data, was equal to 2. The best value of L"(K), the value of the most difference between consecutive levels of K, was at K = 5 (Figure 2). The small improvement for additional K populations is consistent with what was shown in the calculated eigenvector values (Table 1). This reduction in the variation explained by K may have been due more to the limitation of the marker set than to actual lack of differentiation between the lines of the other groups.

The output from STRUCTURE created a visualization of relatedness beyond genetic distance, as percent genetic composition due to the admixture modeling. We considered any entry with greater than 50% relatedness to a single cluster was considered to belong to that cluster. STRUCTURE at K = 2 showed the differentiation of BSSS and NSS lines (Supplementary Figure 2). At K = 3 the B73/Mo17 entries a BSSS x NSS hybrid were identified as a separate group (Supplementary Figure 2). The addition of higher levels of K did not lend to any substantial alteration of the BSSS cluster, but rather identified sub-

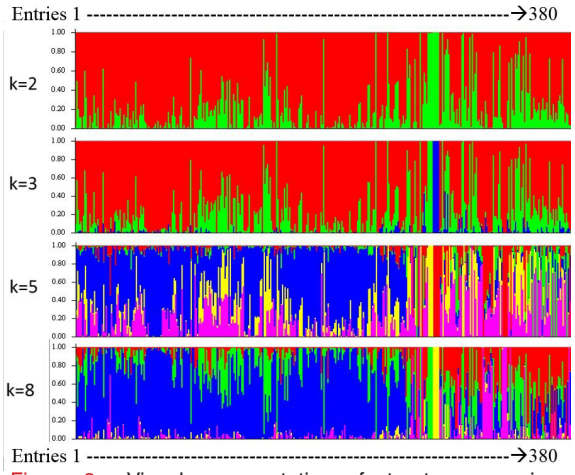

Figure 3 - Visual representation of structure groupings across individuals. Structure uses an admixture model to delegate group designations based on markers present in individuals. Each entry is represented in this graphic by a vertical bar. Entries with background pedigrees from more than one group have more than one color in the vertical section, all group quantifications per individual add to a value of 1. The varying values of K correspond to how many groups Structure forced the entries to fit into. K = 2 and k = 3 are shown in entry numeric order, k = 5 and k = 8 are sorted by cluster association. Colors were assigned to differentiate the populations randomly for each K tested.

heterotic groups within the NSS cluster. Examining the pedigrees in each grouping and from the PCoA, K = 5 made the most logical sense and was used for most subsequent analysis in this study (Supplementary Figure 2). The largest group with K = 5 was composed of the majority of the Texas entries, 222 of the 380 entries, (Figure 3; blue group in lower left quadrant). However, that group did contain other germplasm, for instance PHV63 and PHT60, two Pioneer ex-PVP NSS white lines, both which have acceptable adaptation to Texas as inbred lines. The additional group being elucidated at K = 5 was composed of Iodents and other ex-PVP lines. At this level of K, 44 entries were not classified into any particular group. Analysis of the data successfully validated the findings of the principal coordinate analysis (Figure 1).

Bias of the Genotyping Method

Throughout the analysis of the genotypic data, the ascertainment bias (Frascaroli et al, 2012) of the SNP chip to Midwestern diversity and heterotic groups was prevalent. Analysis of the marker genotype data showed the strongest ability to identify and separate temperate materials, despite them being at lower frequency in the overall population. This demonstrated that the design of the SNP chip was best suited to differentiate for Midwest germplasm, specifically BSSS versus NSS lines, as would be expected based on the design of this resource for elite Midwest germplasm. Of the 635 polymorphic markers, 116 SNPs were identical at greater than 80% of the lines in BSSS and less than 20% in the NSS group, the largest category (Table 2). Although this is a well-defined heterotic pattern, this high focus on the SNPs may have exacerbated it. In contrast, the lines from the Texas breeding program had 27 alleles differentiating the NSS, 8 from SS, and 1 from Iodent (Table 2). Despite this very limited number of population specific alleles, the separation with SS was very strong and the NSS and Iodent's were moderate. It was likely that, instead of unique alleles, it was the combinations of alleles and their frequencies that were useful for differentiating the Texas germplasm from the Midwestern germplasm in STRUCTURE. It stands to reason that this bias in the data could be the reason why the posterior probability analysis did not have strong evidence to identify a true value of K beyond 2. With the PCoA, it would seem the bias was prevalent in the clear first factor of separation of the first principal coordinate, BSSS versus everything else (NSS), but in the second coordinate it was not as clear-cut but became more apparent after the STRUCTURE analysis. If the SNP chip had been tailored to the data set, which has a majority of germplasm from the Texas breeding program, it is likely that a higher level of explanation than 18% could have been achieved (Table 3). It could also be a factor that many crosses were made in the Texas program without regard to heterotic group, as they were temperate x tropical and tropical x tropical instead of temperate x temperate. The bias was seen most predominantly in the STRUCTURE results. In both the PCoA and STRUCTURE the data is separated first into BSSS and everything else (NSS). It was at K = 3 that the ascertainment bias became clear. At K = 3 the Mo17/B73 crosses were identified as a cluster. Since Mo17 was one of the defining members of the NSS cluster and the BSSS are all B73 derived, this would be expected, but shows that the SNP set was developed best differentiate these two inbreds. This was further validated

Table 2 - Summary of marker ability to identify each heterotic group. The left column (>80%) shows the heterotic group for which more than 80% of the lines within that group yield a single allele at a marker. The middle column (count) is how many markers this occurred for. The right column (<20%) are the heterotic groups being compared against the column on the left. The 80% and 20% cutoffs were determined arbitrarily.

>80%	Count	<20%
BSSS	116	NSS
	10	Iodent
	8	Texas
NSS	41	BSSS
	35	Iodent
	25	Texas
Iodent	41	NSS
	10	BSSS
	1	Texas
Texas	27	NSS
	8	BSSS
	1	Iodent

at the K = 4 level when Mo17 and its derivatives were identified as their own cluster. Finally at K = 5 the lines from the Texas program were left in the final cluster, well distinguished from both B73 and Mo17. The two main groups of tropical germplasm used in the Texas breeding program (the LAMAs [see Mayfield et al, 2012 for a review of their background] and the CML from CIMMYT in Mexico) seemingly defined the opposing poles of both the first and second principal coordinate (Supplementary Figure 3), demonstrated by the values below 0.0. The results of those lines clustering so tightly, equally distant from both Mo17 and B73 lends evidence to the fact that they are very dissimilar to either of those pivotal Midwestern lines. The constraints of the markers did not allow for much separation within the larger cluster from the novel Texas group.

Conclusions

This study demonstrated molecular genetic diversity in the Texas program on a macroscopic level. It is clear that these Texas lines are novel from what is otherwise being bred in the Midwestern US. A more refined investigation would likely prove to be more useful for subtropical maize breeders using this Texas material. This would require the development of a SNP array based on lines in the subtropical cluster from the Texas breeding program or the use of next-generation sequencing genotyping such as genotyping by sequencing (Elshire et al, 2011), RAD-seq (Davey et al, 2011), or digital genotyping (Mitra et al, 2003). This would allow for greater differentiation among those lines and the possibility of identifying subheterotic groups within the K = 5 subtropical cluster. The success of the genotyping lent to the mutual validation of the distance matrix calculations and the genetic data. Finding lines of known common pedigrees in close proximity in the analyses confirmed the genotyping data. This gave confidence to the results of other lines and inherently for all results throughout the study. The correlation of genetic distance analysis (PCoA) with the population analysis from STRUCTURE also provided strong evidence for the cogency of the results. There has been substantial focus throughout the US maize research community of the three accepted heterotic groups, BSSS, NSS, and Iodents. Nelson et al (2008) produced an excellent study relating these three groups through a molecular characterization and the results have been a great tool for breeders that use that germplasm. However, the results of Yan et al (2009) among others that have included substantial tropical germplasm suggest that the broad picture of maize may be skewed by a Midwestern focus and that there is likely a greater diversity between tropical and temperate than between the various temperate heterotic groups. Furthermore, much of the tropical material discussed and used in the US are early flowering lines or those from the GEM project that are 50% or 75% elite temperate maize.

Despite ascertainment biases in the markers used, this study confirms the large divide between temperate and tropical germplasm and suggests that the Texas maize breeding program provides a strong contrast to the germplasm selected throughout the rest of the United States. This study also adds resources for breeders working in a subtropical environment, however using genotyping resources designed for temperate material gave results consistent with expectations, but were sub-optimal. The findings in this study has provided evidence at a molecular level to confirm the novelty of the Texas maize breeding program, and that data has provided information well beyond what was already known. The data showed that there was a substantial difference between the germplasm developed at the subtropical research station in Texas and the temperate material from other stations around the US, at a genetic level. When important lines are located on the PCoA graph (Figure 1), the separation becomes clearer. These lines from the Texas program will provide useful sources of diversity for a changing climate in the years to come.

Acknowledgements

The authors would like to thank DuPont Pioneer for in-kind support in genotyping, Dr Russell Jessup for critical comments on an earlier version of this manuscript. SDS was funded by the Willie May Harris Fellowships established by Mr John Richard Harris. Support for this work also came from the USDA-HATCH, the Texas Corn Producers Board, USDA-NIFA-AFRI Plant Breeding and Education Program grant #2010-85117-20539, and USDA-NIFA-AFRI Climate Change Mitigation and Adaptation in Agriculture Program grant #2011-67003-30342.

Table 3 - PCoA analyses were done after the cluster analysis to compare how the PCoA of the main results were affected by the presence of the Texas heterotic group. This shows that the marker set does poorly to identify variation within the Texas group and performs stronger when restricted to temperate materials.

Texas Cluster PCoA			W/O Texas - PCoA		
Coord.	Total	Differ.	Coord.	Total	Differ.
1	5%	5%	1	22%	22%
2	10%	5%	2	32%	10%
3	14%	4%	3	39%	7%
4	17%	3%	4	44%	5%
5	21%	3%	5	48%	4%

References

Alexandratos N, Bruinsma J, 2012. World agriculture towards 2030/2050: the 2012 revision. Food and Agriculture Organization of the United Nations (FAO - 2012)

Balding DJ, Nichols RA, 1995. A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. Genetica 96(1-2): 3-12

Barrero-Farfan, ID, Murray SC, Labar S, Pietsch D, 2013. A multi-environment trial analysis shows slight grain yield improvement in Texas commercial maize. Field Crops Res 149: 167-76

Barrer-Farfan ID, De La Fuente G, Murray S, Isakeit T, 2015. Genome Wide Association Study for Drought, Aflatoxin Resistance, and Important Agronomic Traits of Maize Hybrids in the Sub-Tropics. PLoS ONE 10 (2): e0117737. doi:10.1371/journal.pone.0117737

Betrán FJ, Bockholt A, Fojt F, 2004 Registration of Tx714 maize germplasm line. Crop Sci 44(3): 1028-1028

Buckler ES, Gaut BS, McMullen MD, 2006. Molecular and functional diversity of maize. Curr Opin Plant Biol 9(2): 172-76

Ceccarelli S, Grando S, Maatougui M, Michael M, 2010. Plant breeding and climate changes. J Agr Sci 148: 627-637

Chapman SC, Chakraborty S, Fernanda Dreccer M, 2012. Plant adaptation to climate change— opportunities and priorities in breeding. Crop Pasture Sci 63:251-268

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12(7): 499-510

De Leon C, Pandey S, 1989. Improvement of resistance to ear and stalk rots and agronomic traits in tropical maize gene pools. Crop Sci 29: 12-17

Elshire RJ, Glaubitz JC, Sun Q, 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379

Evanno G, Regnaut S, Goudet J, 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol14(8): 2611-620

Everitt B, 1980. Cluster analysis 2nd ed. New York, NY

Fan J, Yeakley J, Bibikova M, 2004. A Versatile Assay for High-Throughput Gene Expression Profiling on Universal Array Matrices. Genome Res 14: 878885.

Frascaroli E, Schrag TA, Melchinger AE, 2013. Genetic diversity analysis of elite European maize (Zea Mays L) inbred lines using AFLP, SSR, and SNP markers reveals ascertainment bias for a subset of SNPs. Theor Appl Genet 126(1): 133-41

Goodman MM, 1991. Notice of Release of NC296A, NC298 and NC300 maize germplasm lines North Carolina Agricultural Research Service, North Carolina State University, Raleigh, North Carolina

Hallauer AR, Lamkey KR, White PR, 1997. Registration of five inbred lines of maize: B102, B103, B104, B105, and B106 Crop Sci 37(4): 1405-1406

Hayhoe K, VanDorn J, Croley T, Schlegal N, 2010. Regional climate change projections for Chicago and the US Great Lakes. Journal of Great Lakes Research 36:7-21

Holden's Foundation Seeds, Inc. 1991. Corn 'LH195' Plant Variety Protection Certificate Issued May 31 Holden's Foundation Seeds, Inc. 2002. Corn 'LH287' Plant Variety Protection Certificate Issued January 30

Khairallah M, Bohn M, Jiang C, 1998. Molecular mapping of QTL for southwestern corn borer resistance, plant height and flowering in tropical maize. Plant Breeding 117:309-318

Lee EA, Tracy WF, 2009. Modern maize breeding, pp.141-160. In: Handbook of Maize. Springer New York

Liu K, Muse SV, 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21(9): 2128-129

Lu Y, Yan J, Guimarães CT, Taba S, 2009. Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. Theor Appl Genet 120(1): 93-115

Mayfield KL, Murray SC, Rooney WL, Isakeit T, 2011. Confirmation of QTL reducing aflatoxin in maize testcrosses. Crop Sci 51(6): 2489

Mayfield K, Betrán FJ, Isakeit T, Odvody G, 2012. Registration of maize germplasm lines Tx736, Tx739, and Tx740 for reducing preharvest aflatoxin accumulation. Journal of Plant Registrations 6(1): 88

Melchinger AE, Messmer MM, Lee M, 1991. Diversity and relationships among US maize inbreds revealed by restriction fragment length polymorphisms. Crop Sci 31(3): 669

Mikel MA, Dudley JW, 2006. Evolution of North American dent corn from public to proprietary germplasm. Crop Sci 46(3): 1193

Mikel MA, 2008. Genetic diversity and improvement of contemporary proprietary north american dent corn. Crop Sci 48(5): 1686

Mitra RB, Butty VL, Shendure J, Williams BR, 2003. Digital genotyping and haplotying with polymerase colonies. PNAS 100(10): 5926-5931

Munoz P R, Resende MF, Huber DA, 2014. Genomic relationship matrix for correcting pedigree errors in breeding populations: impact on genetic parameters and genomic selection accuracy. Crop Sci 54(3): 1115-1123

Nei M, 1987. Molecular evolutionary genetics Columbia University Press New York Nelson PT, Coles ND, Holland JB, Bubeck DM, 2008. Molecular characterization of maize inbreds with expired US plant variety protection. Crop Sci 48(5): 1673

Poland J, Bradbury P, Buckler E, 2011. Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. PNAS 108(17): 6893-6898

Pollak L, Salhuana W, 2001. The germplasm enhancement of maize (GEM) project: private and public sector collaboration broadening the genetic base of crop production. IPGR/FAO

Pollak L, 2003. The history and success of the public–private project on germplasm enhancement of maize (GEM). Adv Agron 78: 45-87

Pritchard JK, Stephens M, Donnelly P, 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945-959

Ray DK, Mueller ND, West PC, Foley JA, 2013. Yield trends are insufficient to double global crop production by 2050. PLoS ONE 8(6): e66428. doi:10.1371/journal.pone.0066428

Reif, J, Xia X, Melchinger A, 2004. Genetic diversity determined within and among CIMMYT maize populations of tropical, subtropical, and temperate germplasm by ssr markers. Crop Sci 44: 326-334

Rohlf FJ, 1972. An empirical comparison of three ordination techniques in numerical taxonomy. Systematic Zoology 21(3): 271

Rogers JS, Collier JW, 1952. Corn production in texas. Texas Agricultural Experiment Station

Tallury S, Goodman M, 1999. Experimental evaluation of the potential of tropical germplasm for temperate maize improvement. Theoretical and Applied Genetics 98: 54-61

Tarter J, Goodman M, Holland J, 2004. Recovery of exotic alleles in semiexotic maize inbreds derived from crosses between Latin American accessions and a temperate line. Theor Appl Genet 109: 609-617

Yan L, Yan J, Guimaraes C, 2009. Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. Theor Appl Genet 120: 93-115

Yu J, Pressoir G, Briggs WH, 2005. The comparisons with diversity that exists in industry, expired plant variety protected (ex-PVP; Mikel and Dudley 2006) lines, remain unknown. Nat Genet 38(2): 203-208