

RESEARCH ARTICLE

Genetic relatedness of previously Plant-Variety-Protected commercial maize inbreds

Travis J. Beckett^{1*}, A. Jason Morales², Klaus L. Koehler², Torbert R. Rocheford¹

1 Department of Agronomy, Purdue University, West Lafayette, IN, United States of America, **2** Dow AgroSciences, Indianapolis, IN, United States of America

* travisbeckett@purdue.edu



Abstract

The emergence of high-throughput, high-density genotyping methods combined with increasingly powerful computing systems has created opportunities to further discover and exploit the genes controlling agronomic performance in elite maize breeding populations. Understanding the genetic basis of population structure in an elite set of materials is an essential step in this genetic discovery process. This paper presents a genotype-based population analysis of all maize inbreds whose Plant Variety Protection certificates had expired as of the end of 2013 (283 inbreds) as well as 66 public founder inbreds. The results provide accurate population structure information and allow for important inferences in context of the historical development of North American elite commercial maize germplasm. Genotypic data was obtained via genotyping-by-sequencing on 349 inbreds. After filtering for missing data, 77,314 high-quality markers remained. The remaining missing data (average per individual was 6.22 percent) was fully imputed at an accuracy of 83 percent. Calculation of linkage disequilibrium revealed that the average r^2 of 0.20 occurs at approximately 1.1 Kb. Results of population genetics analyses agree with previously published studies that divide North American maize germplasm into three heterotic groups: Stiff Stalk, Non-Stiff Stalk, and Iodent. Principal component analysis shows that population differentiation is indeed very complex and present at many levels, yet confirms that division into three main sub-groups is optimal for population description. Clustering based on Nei's genetic distance provides an additional empirical representation of the three main heterotic groups. Overall fixation index (F_{ST}), indicating the degree of genetic divergence between the three main heterotic groups, was 0.1361. Understanding the genetic relationships and population differentiation of elite germplasm may help breeders to maintain and potentially increase the rate of genetic gain, resulting in higher overall agronomic performance.

OPEN ACCESS

Citation: Beckett TJ, Morales AJ, Koehler KL, Rocheford TR (2017) Genetic relatedness of previously Plant-Variety-Protected commercial maize inbreds. PLoS ONE 12(12): e0189277. <https://doi.org/10.1371/journal.pone.0189277>

Editor: Aimin Zhang, Institute of Genetics and Developmental Biology Chinese Academy of Sciences, CHINA

Received: July 17, 2017

Accepted: November 23, 2017

Published: December 13, 2017

Copyright: © 2017 Beckett et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All genotype-by-sequencing (GBS) data are available from the www.panzea.org database. As the samples are part of a larger data set, individual accession identifiers for all inbreds are listed S4 Table. The genotype for each individual may be obtained by pasting the "Taxa" identifier in the appropriate place on this online database webpage: <http://cbsuss05.tc.cornell.edu/hdf5new/query.asp>.

Funding: This research was funded by a grant from Dow AgroSciences (www.dowagro.com). The

Introduction

Maize (*Zea mays* subsp. *mays*) is one of the most important agricultural crops in the United States. Grain yields remained generally constant from 1866 to 1936, when the vast majority of

funder provided additional support in the form of salaries for authors AJM and KLK, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section.

Competing interests: We have read the journal's policy and the authors of this manuscript have the following competing interests: Authors KLK and AJM are employed by Dow AgroSciences. This commercial affiliation does not alter our adherence to PLOS ONE policies on sharing data and materials.

maize was grown from open-pollinated seeds (see Fig 1). Foundational studies in the early 1900's on inbreeding and heterosis introduced the idea of producing commercial maize seed on a hybrid plant resulting from a cross of two inbreds [1–8]. Subsequently, the replacement of open-pollinated varieties with double- and single-cross hybrids played a major role in sustained increases in grain yield since 1937 [8, 9].

After over three decades of widespread commercial hybrid maize production, the Plant Variety Protection Act (PVPA) was passed by the U.S. Congress in 1970 [11]. This law guaranteed intellectual property rights to developers of new plant varieties by prohibiting others from reproducing, selling, or importing any protected variety, for a period of 18 (presently 20) years [12]. New plant varieties may also be protected by U.S. patents. The legality of granting patents for plants was affirmed by rulings by the U.S. Supreme Court in *Diamond v. Chakrabarty* (1980) [13] and *J.E.M. Ag Supply v. Pioneer* (2001) [14], and by the U.S. Board of Patent Appeals and Inferences in *Ex parte Hibbert* (1985) [15–17]. Both utility patents and PVP certificates are effective forms of germplasm protection commonly used by U.S. private-sector soybean and maize breeders [18, 19].

When a PVP certificate issued for a maize inbred expires, and there is no active patent protecting the property, the inbred then enters the public domain and is provided free of charge

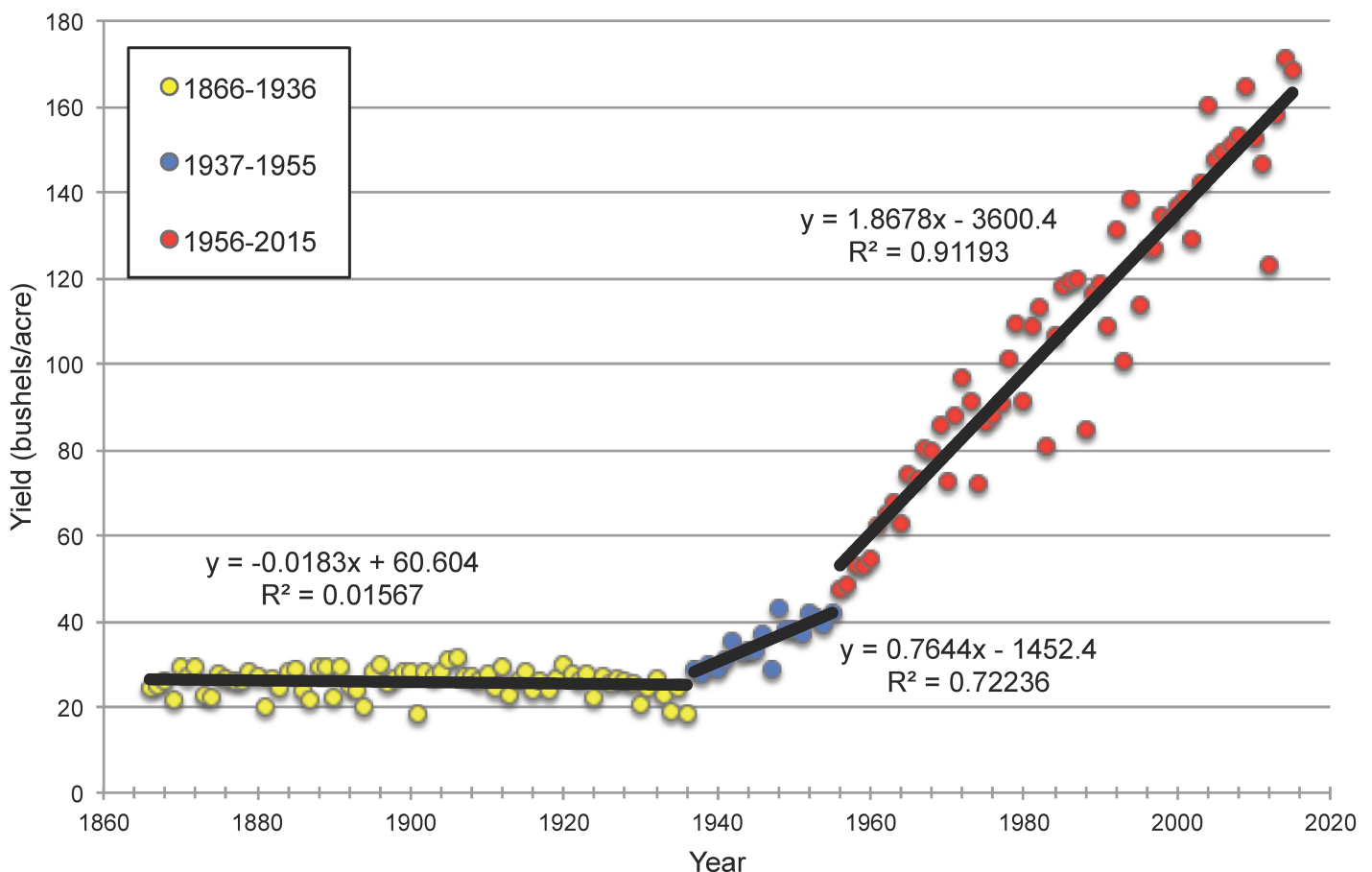


Fig 1. Historical U.S. Maize Yields, 1866 to 2015. Data is separated into three time periods according to the source of corn seed planted for agricultural production. In the first period, from 1866 to 1936, the vast majority of corn grown was of the open-pollinated type. During the second period, from 1937 to 1955, most hybrid corn planted in the U.S. was produced from double crosses. Throughout the third period, from 1956 to 2015, single-cross hybrids were the largest source of corn seed planted for commercial production. A best-fit linear trend is included for each time period. Data was obtained from the USDA National Agricultural Statistical Service [10].

<https://doi.org/10.1371/journal.pone.0189277.g001>

by the United States Department of Agriculture (USDA). Many of these now-publicly available inbred lines have made a significant contribution to current commercial germplasm. By pedigree analysis, Mikel (2011) found that the top four progenitors (by genetic contribution) of 305 maize inbreds registered by PVP and/or utility patent between the years 2004 and 2008 were 3IIH6 (12.2%), B73 (11.7%), PH207 (9.5%), and PHG29 (9.4%) [20]. Three of these inbreds, 3IIH6, PH207, and PHG29, have expired PVP certificates. All four inbreds mentioned above are included found the population used in this study. Each inbred with a newly expired PVP certificate is a readily available source of highly selected alleles and haplotype blocks that can likely improve germplasm pools in breeding programs that previously did not have access to such elite genetics [20].

Heterosis, also known as hybrid vigor, is observed when the F1 progeny of a cross between two individuals from different germplasm groups performs better than the F1 progeny of a cross between two individuals from the same germplasm group [21]. In maize, these germplasm groups are generally referred to as heterotic groups. Numerous proposals of North American maize heterotic group divisions have been made [21–26]. Most are some variant of the dominant heterotic pattern of Stiff Stalk (SS) and Non-Stiff Stalk (NSS), commonly known as the female-male heterotic pattern. A representative summary of the heterotic group proposals is given in Table 1.

Maize inbreds can be assigned to heterotic groups based on: (1) pedigree information and specific combining ability based on field trials; (2) molecular markers and genetic relatedness analysis; or (3) some combination of these two methods [27]. Many attempts to classify public maize lines into heterotic groups using molecular markers have been reported, with varying levels of success (See Table 2). Early molecular marker platforms produced small number of markers at inconsistent accuracy levels [28–34]. One problem with using a small number of markers is that it can be difficult to precisely resolve the heterotic and family group membership of closely related inbred lines, as the marker set may not include all loci that are responsible for heterotic divergence. Consequently, genetic-based determination of heterotic groups and combining ability was not considered as effective as traditional field-validation at accurately identifying similar groups of germplasm out of a large group of seemingly unrelated inbred lines [29, 34]. Genotyping technology has now improved to the point where genotype-based heterotic groupings appear just as accurate as the groupings defined by pedigrees and empirical field measures of combining ability [37, 40]. Next-generation sequencing methods such as genotyping-by-sequencing (GBS) can be very helpful in determining the heterotic group position of newly released ex-PVP inbreds relative to a breeding program’s existing inbreds.

Table 1. Summary of proposed heterotic group divisions in maize.

Author(s)	Year	Proposed heterotic group division
Troyer [22]	1999	Reid Yellow Dent, Minnesota 13, Northwestern Dent, Lancaster Sure Crop, and Learning Corn
Gethi, et al. [23]	2002	Reid Yellow Dent, Lancaster, and other sub-groups
Mikel, et al. [24]	2006	Seven key ancestral inbreds: B73, Mo17, PH207, PHG39 (from B37), LH123Ht, LH82, and PH595
Nelson, et al. [25]	2008	B73, Mo17, PH207, A632, Oh43, B37, and mixed
Lu, et. al. [26]	2009	Iowa Stiff-Stalk Synthetic (BSSS) and non-Iowa Stiff-Stalk Synthetic (non-BSSS)
Bernardo [21]	2014	BSSS (B14, B37, B73) and non-BSSS (Iodent, Oh43, Mo17, and other subgroups)

<https://doi.org/10.1371/journal.pone.0189277.t001>

Table 2. Using molecular markers to identify heterotic groups in maize.

Author(s) and year	No. of Inbreds	Population description	—Markers—	
			No.	Type
Smith, et al. (1990) [28]	37	North American temperate	257	RFLP
Dudley, et al. (1991) [29]	14	North American temperate	52	RFLP
Livini, et al. (1991) [30]	40	Italian temperate	149	RFLP
Melchinger, et al. (1991) [31]	32	North American temperate	83	RFLP
Mumm & Dudley (1994) [32]	148	North American temperate	46	RFLP
Senior, et al. (1998) [33]	94	North American temperate	70	SSR
Barata & Carena (2006) [34]	40	North American temperate	49	SSR
Nelson, et al. (2008) [25]	109	North American temperate ^a	614	SNP
Lu, et al. (2009) [26]	770	Tropical and temperate from CIMMYT, China, and Brazil	449	SNP
Kahler, et al. (2010) [37]	98	North American temperate ^a	285	SSR
van Heerwaarden, et al. (2012) [38]	294	North American temperate ^a	45,997	SNP
Olmos et al. (2013) [35]	103	Argentinean temperate	50	SSR
Romay et al. (2013) [39]	2,185	Global tropical and temperate ^a	681,257	SNP
Unterseer, et al. (2014) [40]	315	Global tropical and temperate	609,442	SNP
Smith, et al. (2015) [36]	380	North American temperate and sub-tropical ^a	635	SNP
Wu, et al. (2016) [41]	544	CIMMYT inbreds	362,008	SNP
Zhang, et al. (2016) [42]	362	Chinese tropical and temperate	56,110	SNP

^aIncludes inbreds with expired Plant Variety Protection certificates.

<https://doi.org/10.1371/journal.pone.0189277.t002>

There are some challenges, however, presented by the GBS method. The success of GBS depends on a minimum read depth, or number of repeated sequences covering a specific locus. Read depth can vary across the genome, between separate GBS batches, and even between individuals [43]. Due to low coverage of sequencing, there may be large portions of the genome without any successful marker calls [44]. Therefore, each set of GBS data—and even each individual genotype—has a unique distribution of the number and quality of genotype calls. Fortunately, when missing data remains after filtering, it can usually be imputed at acceptable levels of accuracy—a very cost-favorable alternative to sequencing at a higher depth [43, 45, 46].

Following the development of next-generation sequencing platforms, there have been a number of studies published on genetic classification of maize inbred lines [25, 26, 36–42]. Several included inbreds with expired Plant Variety Protection certificates: Nelson, et al. (2008) with 92 ex-PVP inbreds; Kahler, et al. (2010) with 33; van Heerwaarden, et al. (2012) with 137; Romay, et al. (2015), with 212; and Smith, et al. (2015) with 105. Out of these, the publication most closely aligned to the subject of this study is that authored by Romay et al. (2013) [38].

This paper presents a comprehensive genotype-based population analysis of all ex-PVP maize inbreds available as of the end of 2012. The robust array of analyses includes measures of genetic diversity, linkage disequilibrium, genotypic clustering, and heterotic groupings. Included in this study is a greater number of ex-PVP inbreds (283) and a wider range of analytical methods than found in previous publications. The results herein can help maize breeders determine how to best incorporate the ex-PVP inbreds into their existing germ-plasm pools.

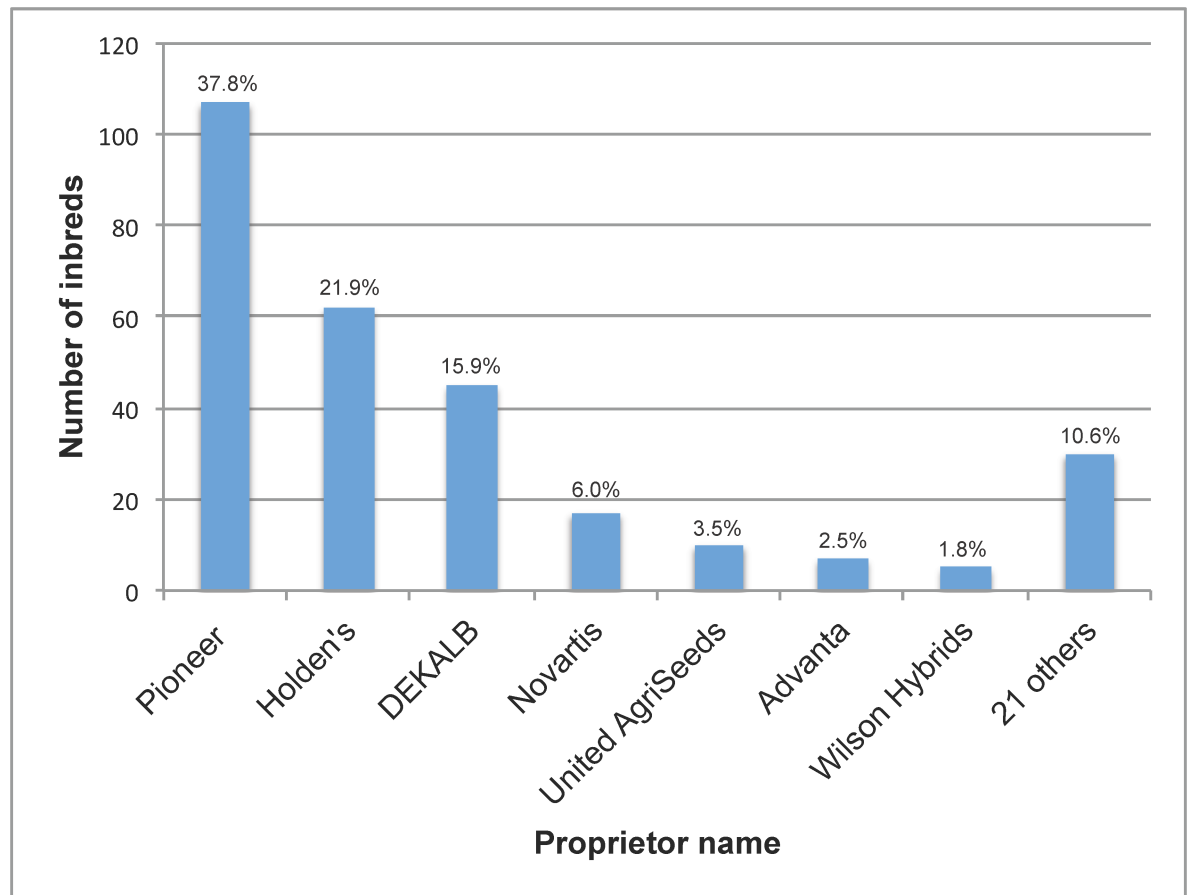


Fig 2. Plant Variety Protection certificates expired as of 2012, by proprietor. Proprietor names were abbreviated as follows: Pioneer, Pioneer Hi-Bred International, Inc.; Holden's, Holden's Foundation Seeds, Inc.; DEKALB, DEKALB Genetics; Novartis, Novartis Seeds, Inc; United AgriSeeds, United AgriSeeds, Inc.; Advanta, Advanta Technology Limited; and Wilson Hybrids, Wilson Hybrids Inc. Proprietor names are on the x-axis, and the number of inbreds present in this set of 283 is on the y-axis. Above each bar is the value in percent, calculated as the number of PVP inbreds for each respective proprietor divided by 283. Proprietorship was obtained from the Plant Variety Protection certificate for each inbred, accessible in the United States Department of Agriculture Agricultural Research Service Germplasm Network Information Database [48].

<https://doi.org/10.1371/journal.pone.0189277.g002>

Materials and methods

Plant material

The maize varieties used in this study include 283 ex-PVP inbreds and 66 public inbred founders. The 283 ex-PVP inbreds were those with certificates that had expired between 1994 and 2012. Distribution by proprietor of these 283 inbreds with expired Plant Variety Protection (ex-PVP) is shown in Fig 2. Pedigrees of the 283 ex-PVP lines were examined and 66 public founder inbreds were identified based on two criteria: (1) the public inbred appeared in the pedigree of at least one ex-PVP inbred; and (2) seed for that public line was available at the start of this project [47]. Seed for all 349 inbreds was requested from the USDA-ARS National Genetic Resources Program [48], and received from the USDA-ARS North-Central Regional Plant Introduction Station (NCRPIS) in Ames, Iowa. Ex-PVP inbred pedigrees were obtained from the PVP certificates, accessed at ars.grin.gov [48]. Public inbred pedigrees were obtained from the volume titled, Compilation of North American Maize Breeding Germplasm [49].

Tables with general information about both the ex-PVP and public founder inbred sets are provided in the supplementary information (see [S1](#) and [S2](#) Tables).

A bar chart showing the distribution of the 283 ex-PVP inbreds used in this study, sorted by proprietor, is displayed in [Fig 2](#). Pioneer Hi-Bred International, Inc. (Pioneer) produced the most inbreds, with nearly 40 percent of these PVP certificates. The top three proprietors, Pioneer, Holden's Foundation Seeds, and DEKALB Plant Genetics, together held over 75 percent of PVP certificates for inbred lines used in this data set. The top seven proprietors, which also includes Novartis Seeds, Inc., United AgriSeeds, Inc., Advanta Technology Limited, and Wilson Hybrids, Inc., accounted for nearly 90 percent of PVP certificates. The remaining ex-PVP inbreds used in this study originated from twenty-one different companies, with between one and three certificates held by each company. Thus, of the North American commercial maize inbreds with PVP certificates that had expired as of the end of 2012, the vast majority (nearly 90%) came from only one-quarter of all private maize breeding programs that used PVP for their inbreds (seven out of twenty-eight companies).

Genotypic data compilation

The original genotypic data comes from two sources. The first source includes genotyping data on 224 lines whose PVP certificates had expired as of the end of 2011, as well as 67 public founder inbred lines. Partially imputed GBS data for these 291 lines was downloaded from the online GBS data repository at www.panzea.org [50]. The build version was ZeaGBSv27, with 955,690 SNPs on AGPv2 coordinates, produced using the enzyme ApeKI and the protocol described by Elshire et al., (2011) [43, 51]. The second source consisted of GBS data on 58 additional ex-PVP inbred lines whose PVP certificates expired during the first four months of 2012. These 58 lines were grown at the Purdue Agronomy Center for Research and Education (ACRE) in West Lafayette, Indiana, in the summer of 2012. Tissue sampling and DNA extraction was performed according to the protocol of Romay et al., (2013) [38]. The DNA samples were sent to the Cornell University Institute for Genomic Diversity (Ithaca, New York), where GBS libraries were prepared and analyzed according to Elshire et al. (2011) [43], using the enzyme ApeKI for digestion and creating a library with 240,021,078 unique barcodes. The GBS pipeline for these 58 lines resulted in 546,531 unfiltered SNPs. The two genotypic data sets were aligned and merged, using TASSEL 5.0, version 20151210 [52]. A summary of the genotypic data set compilation steps is given in [Table 3](#).

Data analysis

SNP characteristics. Quality control measures were employed to ensure that the genotypic data would be as accurate as possible for population structure analysis. Genotypic markers with missing data greater than ten percent and/or a minor allele frequency (MAF) less than 0.05 were removed. As the genotypic analyses assume only two alleles per locus, minor SNP statuses (i.e. tertiary and greater alleles) were changed to missing data. Additionally, any heterozygote calls were changed to missing data. Applying these filters reduced the maximum amount of missing data per inbred to no more than 30 percent for any one inbred in this data set (see [Table 4](#)). The specific level of 30 percent was chosen to balance the share of missing data between the two previously unmerged GBS sets while also minimizing the proportion of overall missing data, thus reducing overall proportion of genotypic errors caused by imputation [53]. These filter thresholds left the genotypic data set with a total of 77,314 SNPs.

The 6.22 percent of genotypic data points that remained as missing data were fully imputed using the 'markov' function in the package 'NAM' in RStudio version 0.98.1103 [53–55]. This function employs a Hidden Markov Model (HMM); however, unlike other HMM-based

Table 3. Summary statistics of unmerged genotypic data sets, before filtering and imputing.

Statistic	GBS set no. 1 ^a	GBS set no. 2 ^b	Merged GBS set
Inbreds	291	58	349
Sites	955,690	546,531	1,281,671
Total data points ^c	278,105,790	31,698,798	447,303,179
Missing SNPs	36,217,717	13,724,629	187,442,397
Percent missing	13.0%	43.3%	41.9%
No. Heterozygous	502,897	596,243	1,100,366
Percentage het.	0.18%	1.88%	0.25%
	Sites common to both GBS sets		220,550
	Percent Sites common to both GBS sets ^d		17.2%

^a The data for these 291 inbreds was obtained from from the online GBS data repository at www.panzea.org [50].

^b The GBS data set for these 58 inbreds was produced by Cornell University Institute for Genomic Diversity (Ithaca, New York).

^c Total data points = Inbreds x Sites = Total number of SNP reads.

^d Calculated by (Sites common to both GBS sets) / (Merged Sites)

<https://doi.org/10.1371/journal.pone.0189277.t003>

imputation methods, the ‘markov’ function only runs in the forward direction and not the backward direction. This feature enables quicker imputation computations for very large data sets. Imputation accuracy was calculated by comparing a completed genotypic data set with a version of the same data set which included imputed values at randomly placed missing data points. Calculations to assess imputation accuracy were repeated 100 times using the same complete data set, with the average amount of randomly placed missing data across the repetitions set at 6.22 percent. The mean imputation accuracy of these repetitions was reported as the overall imputation accuracy for this data set.

Principal component analysis. Principal component analysis (PCA, or PC analysis) was performed by the ‘prcomp’ function in RStudio [55]. The optimal number of PCAs was determined by consulting both the scree plot and the PCA plots, in context of what has already been reported about the number of major maize heterotic group divisions [21–26].

Linkage disequilibrium. Analysis of linkage disequilibrium (LD) was performed in RStudio [55] with the package ‘NAM’, using the function ‘ld’ [54]. Decay of LD was determined for

Table 4. Summary statistics of merged, filtered, and imputed genotypic data sets.

Statistic	Merged data ^a	Filtered data ^b	Imputed data ^c
Inbreds	349	349	349
Sites	1,281,671	77,314	77,314
Total data points ^c	447,303,179	26,987,123	26,987,123
Missing SNPs	187,442,397	1,680,218	0
Percent missing	41.9%	6.22% ^d	0
Heterozygous	1,100,366	0	0
Percentage het.	0.25%	0	0

^a Two genotyping-by-sequencing (GBS) data sets were merged for this study. One consisted of 546,531 SNP reads on 58 inbreds, and the other had 1,290,050 SNP reads on 291 inbreds.

^b Filtering consisted of: (1) removing markers with minor allele frequency (MAF) less than 0.05; (2) removing markers with greater than 17.2 percent missing data; and (3) changing the genotype call to missing at all heterozygous sites and all minor SNP states (tertiary and above).

^c Total data points = Inbreds x Sites = Total number of SNP reads.

^d For missing SNPs per inbred: median was 3.90%; maximum was 28.5%; and minimum was 0.66%.

<https://doi.org/10.1371/journal.pone.0189277.t004>

each chromosome individually by considering all pairwise SNP marker combinations. For each SNP pair, both the distance (bp) and the coefficient of determination r^2 were calculated, then plotted. A smoothing function within RStudio ('lokern') was employed to insert a trend line for each chromosome [56]. A trend line for the mean LD over all chromosomes was also included in the plot.

Population structure. Population substructure was analyzed using RStudio [55], using various packages as described below. Nei's distance, calculated by

$$D = -\ln \frac{\sum_l \sum_u X_u Y_u}{\sqrt{(\sum_l \sum_u X_u^2)(\sum_l \sum_u Y_u^2)}} \quad (1)$$

was used to create the distance matrix with functions called from the package 'NAM' [57]. The built-in R function 'hclust' [55] was used to perform an hierarchical cluster analysis using Ward's minimum-variance method [58], defined by

$$d_{ij} = d(\{X_i\}, \{X_j\}) = \|X_i - X_j\|^2 \quad (2)$$

A genetic clustering diagram, a dendrogram, was created and coded using the package 'ape' [59]. The tree was exported in Newick (also known as New Hampshire) file format, then imported into the online application Interactive Tree of Life (iTol) for color annotating [60].

Once the tree was created, the number of sub-groups was determined by a multi-step approach. First, the plots produced from principal component analysis were examined for indications of separation into clear groups. The function 'cutree' in RStudio [55] was then used to split the tree into sub-groups based on branch length (genetic dissimilarity), informed generally by the number of clear groups indicated by the principal component plot. Known pedigrees and results of previous studies [25, 36, 47] were then used to identify group names and confirm boundaries. Divisions of between three and eight sub-groups were examined in more detail. Maximum sub-groups were reached when further division did not appear justified based on pedigrees, results of previous studies, as well as the principal component analysis.

Genetic diversity

To assess the level of genetic variation when dividing the population into three main heterotic groups of SS, NSS, and Iodent, F_{ST} was calculated using the package 'NAM' [54] in RStudio [55]. This analysis produces estimates of unbiased F_{ST} statistics by a weighted analysis of variance method [61]. Overall F_{ST} was calculated as the simple average across all loci.

To reduce bias in the F_{ST} statistic, two important interrelated modifications were made [62]. Both involved filtering of the inbreds to be used in calculation of the F_{ST} statistic. First, to correct for sample size among sub-populations (or heterotic groups, in this case), a balanced number of individuals across the three heterotic groups was selected. Second, to reduce bias of allelic frequencies caused by pedigree structure, the balanced sample from each heterotic group was composed of individuals as genetically unrelated as possible. For example, within the Stiff Stalk heterotic group, the inbreds F42 and B73 are very genetically closely related. Including both of these inbreds in an allelic frequency measure would be essentially using duplicate genotypic data, and would bias the allele frequency calculated for the Stiff Stalk heterotic group. Simply excluding either one inbred, however, while retaining the other, removes the pedigree structure bias while retaining sufficient genetic diversity in the context of F_{ST} analysis.

The Iodent heterotic group contained the least number of individuals, so the filtering process was initiated within this subgroup. Filtering of the Iodent subgroup according to the two

criteria described above resulted in 44 remaining inbreds. Therefore, in order to balance the data set with equal number of individuals from each heterotic group, 44 became the target number of individuals to select out of the remaining two groups. A list of the inbreds selected for F_{ST} analysis is included in the supplemental materials in [S3 Table](#).

For the SNPs with the highest F_{ST} values, a candidate gene search was completed for a 10 kbp window on either side of the SNP. This candidate gene search was done within the B73 v2 reference genome, using the R package ‘Zbrowse’ [63].

Results

Marker coverage and missing data

As the genotypic data set came from two different sources, it was necessary to merge the genotypic data before analysis. Consequently, out of a total of 955,690 SNP markers in the first set and 546,531 SNP markers in the second set, only 220,550 sites—or 17.2 percent—were common to both GBS sets ([Table 3](#)). Following the GBS data set merger and then filtering to remove heterozygous calls, monomorphic sites, markers with greater than 17.2 percent missing data, and SNPs with minor allele frequency (MAF) less than 0.05, the number of SNP markers remaining was 77,314 ([Table 4](#)). Missing data may not have been distributed randomly, as use of the B73 reference genome for read alignment causes inbreds closely related to B73 to have a lower proportion of missing data than inbreds more distantly related to B73 [38]. Even so, prior to population analysis, missing data was reduced to zero by imputation. Imputation accuracy was estimated to be 0.83.

Linkage disequilibrium

[Fig 3](#) shows the decay of linkage disequilibrium (LD) across genetic distance. An average LD of r^2 at 0.2 was reached at approximately 1.1 Kbp. All chromosomes followed the same general decay trend, with the exception of chromosome 7, which appeared to decay more rapidly than the rest between 100 bp and 1 Kb, reaching an average LD of 0.2 at approximately 1 Kb.

Population structure

The ex-PVP inbreds originated from 28 different proprietors ([Fig 2](#) and [S1 Table](#)). The public founder inbreds originated from research programs located in 17 different states and one Canadian province ([S2 Table](#)). Population stratification was expected to follow the three principal heterotic groups of maize: Stiff Stalk, Non-Stiff Stalk, and Iodent. Two dimensional PC analysis validated this expectation ([Fig 4](#)), with three clear spatial divisions in the PCA plot corresponding with the three main population groups identified in the phylogenetic cluster analysis. A PCA plot with three principal components for each inbred line also shows a clear division into three main groups ([Fig 5](#)).

Further confirmation of the generally expected population stratification is visible in the scree plot of the principal component analysis ([Fig 6](#)). The optimal number of principal components to explain genotypic variation, three, was found by visually determining the largest point of inflection, or “elbow” of the non-linear trend line [64]. To find the optimal number of principal components, more complex and empirical methods—such as the silhouette method [65] or the Gap statistic [66]—could have been employed. However, in context of prior knowledge of North American maize heterotic groups as well as phylogenetic cluster analysis based on genotypic data (see next paragraph), the “elbow” method is more than sufficient in this case. Percent variation explained by additional principal components is depicted in [Fig 7](#).

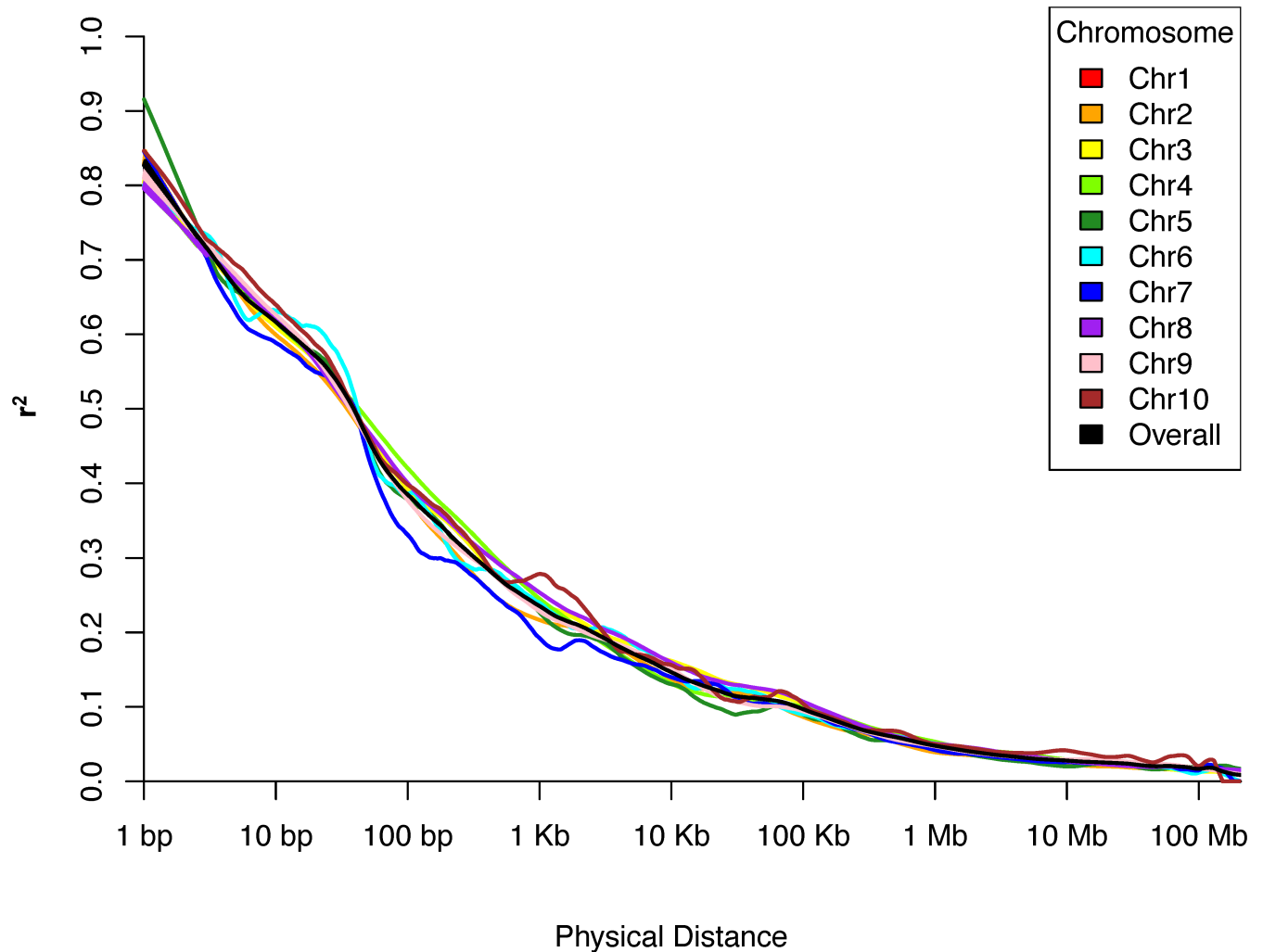


Fig 3. Decay of linkage disequilibrium with physical distance. Decay of linkage disequilibrium (LD) with physical distance between 77,314 pairs of single nucleotide polymorphism (SNP) markers in the ex-PVP and public founder genotypic data set. Physical distance (scaled logarithmically) is on the x-axis and LD, measured in r^2 is on the y-axis. Individual chromosomes are indicated by line color, with the overall average of all data overlaid as a black trend line.

<https://doi.org/10.1371/journal.pone.0189277.g003>

Phylogenetic cluster analysis produced a dendrogram that divided into three main groups, Stiff Stalk, Non-Stiff Stalk, and Iodent (Fig 8). General heterotic group assignments based on pedigree data as well as previous publications agree with the classifications assigned by the genotypic clustering method used herein [20, 25, 29, 35, 36, 47]. For a more detailed examination of heterotic group classifications, a dendrogram divided into eight principal population sub-groups was produced by the same methods of cluster analysis. This dendrogram with eight divisions is included in the supplementary materials (see S1 Fig).

Genetic diversity

The overall genetic diversity, or mean F_{ST} , when considering divisions into SS, NSS, and Iodent heterotic groups is 0.1732. Genome-wide F_{ST} values plotted against relative marker position are presented in Fig 9. This plot reveals trends that may be worth further study, as

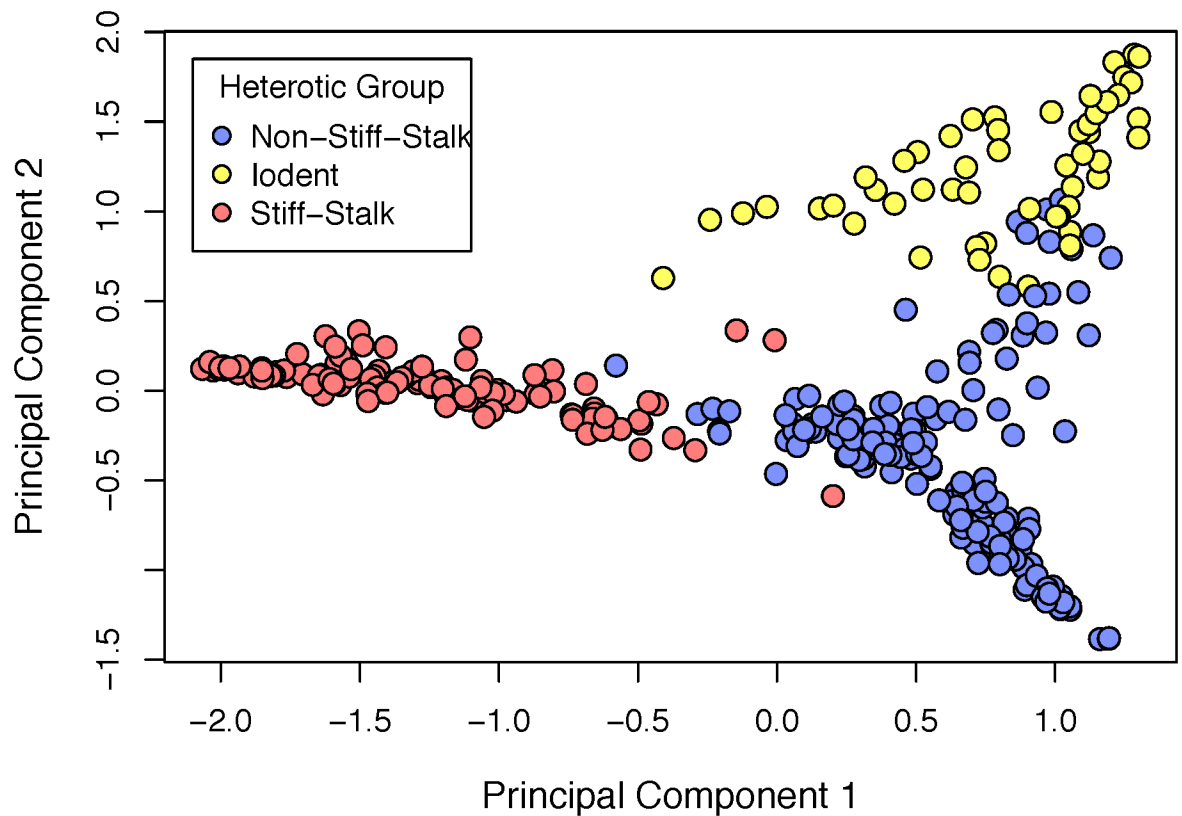


Fig 4. Principal component 1 vs. principal component 2. Principal component no. 1 (x-axis) vs. principal component no. 2 (y-axis), color annotated by three heterotic group divisions. Colors indicate membership in one of three population sub-groups as determined by phylogenetic cluster analysis.

<https://doi.org/10.1371/journal.pone.0189277.g004>

they may be indicative of individual loci and/or genomic regions that are involved in population differentiation or possibly even heterosis.

Individual loci were examined for proximity to candidate genes; results are presented in Table 5. These 11 loci are representative of areas where the highest F_{ST} values for individual SNPs were found, specifically in chromosomes 2, 4, 5, 7, and 8. The four genomic regions with the highest mean F_{ST} value over a window of 70 SNPs are identified in Table 6.

Discussion

Linkage disequilibrium

The rate of LD decay reported in maize depends on the population, the genetic region(s) under study, as well as the statistical methods used to compute the values. For the population in this study, the result of LD decay (r^2 at 0.2 at approximately 1.1 Kbp) appears reasonable and well within the general range reported in previous studies. As LD values reported in the literature can vary, the following summary of relevant results of LD decay provides context within which the results of this study can be evaluated.

Developments in genotyping methods led to many studies in the early 2000's that reported on LD in maize. Tenaillon et al., (2001) [68], reported on a set of 25 maize genotypes comprised of 16 landraces and nine North American inbred lines, analyzing the LD between 21 SNP markers on chromosome 1. Their results showed that LD declines to an average r^2 of 0.2 at 300 bp in the mixed set of 25 genotypes, and greater than 1 kbp in the subset that includes

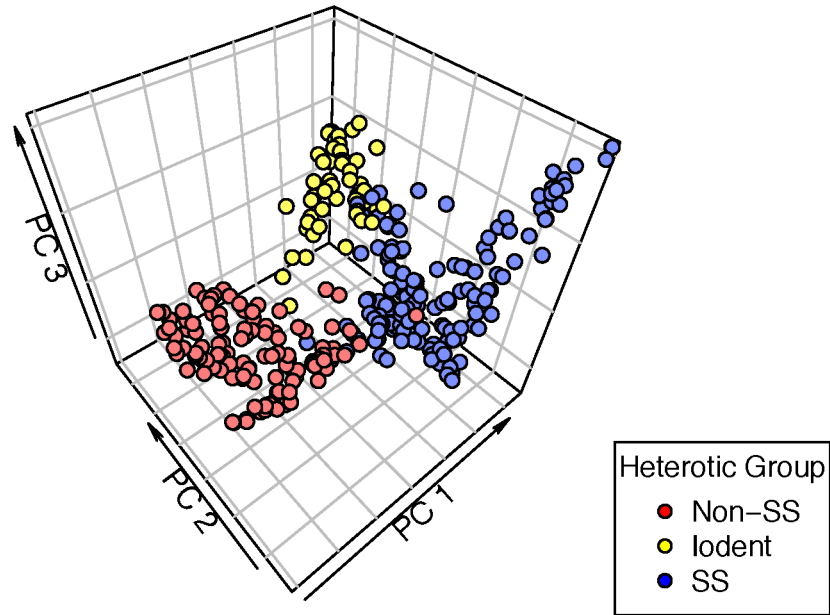


Fig 5. Three-dimensional plot of principal component analysis. Axes labels are abbreviated for principal components 1, 2, and 3, respectively. Colors indicate membership in one of three population sub-groups as determined by phylogenetic cluster analysis.

<https://doi.org/10.1371/journal.pone.0189277.g005>

only the nine North American inbreds. Remington et al., (2001) [69] analyzed the LD for 102 inbred lines from a broad range of temperate and tropical origins, using over 1.5 Mb of SSR marker data centered on six candidate gene regions, and found that the decay reached an average of average r^2 of 0.2 at an average of 550 bp for five of the six candidate regions. For the

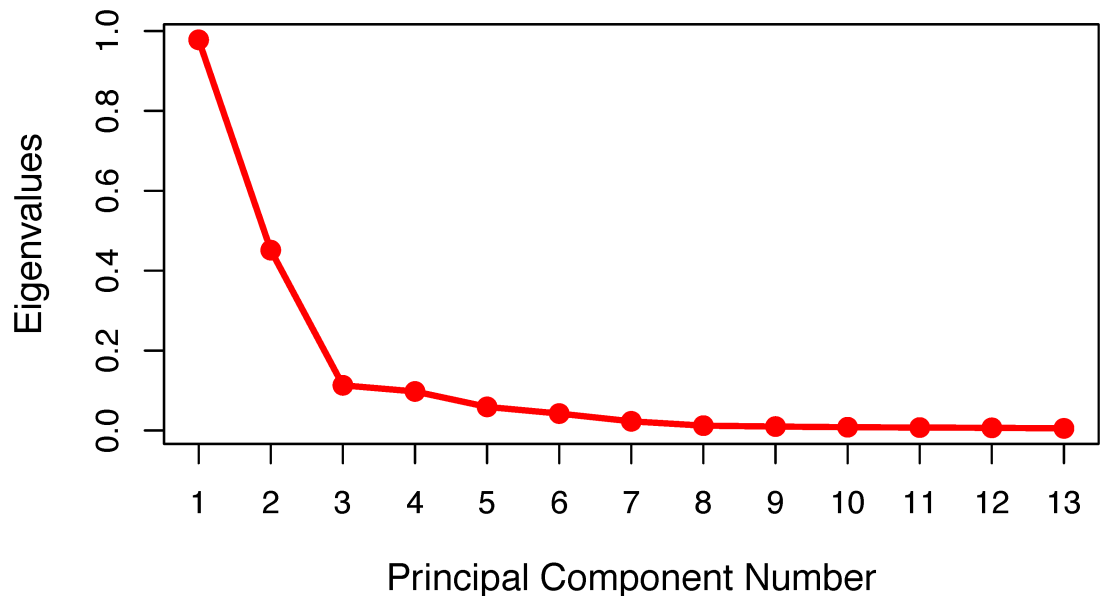


Fig 6. Scree plot of principal component analysis. The number of principal components (PCs) is on the x-axis and the associated eigenvalues—which indicate the amount of variance yet unexplained—are on the y-axis. The optimal number of principal components to explain the variation found in the genotype is found by visually determining the largest point of inflection, or “elbow” of the non-linear trend line [64].

<https://doi.org/10.1371/journal.pone.0189277.g006>

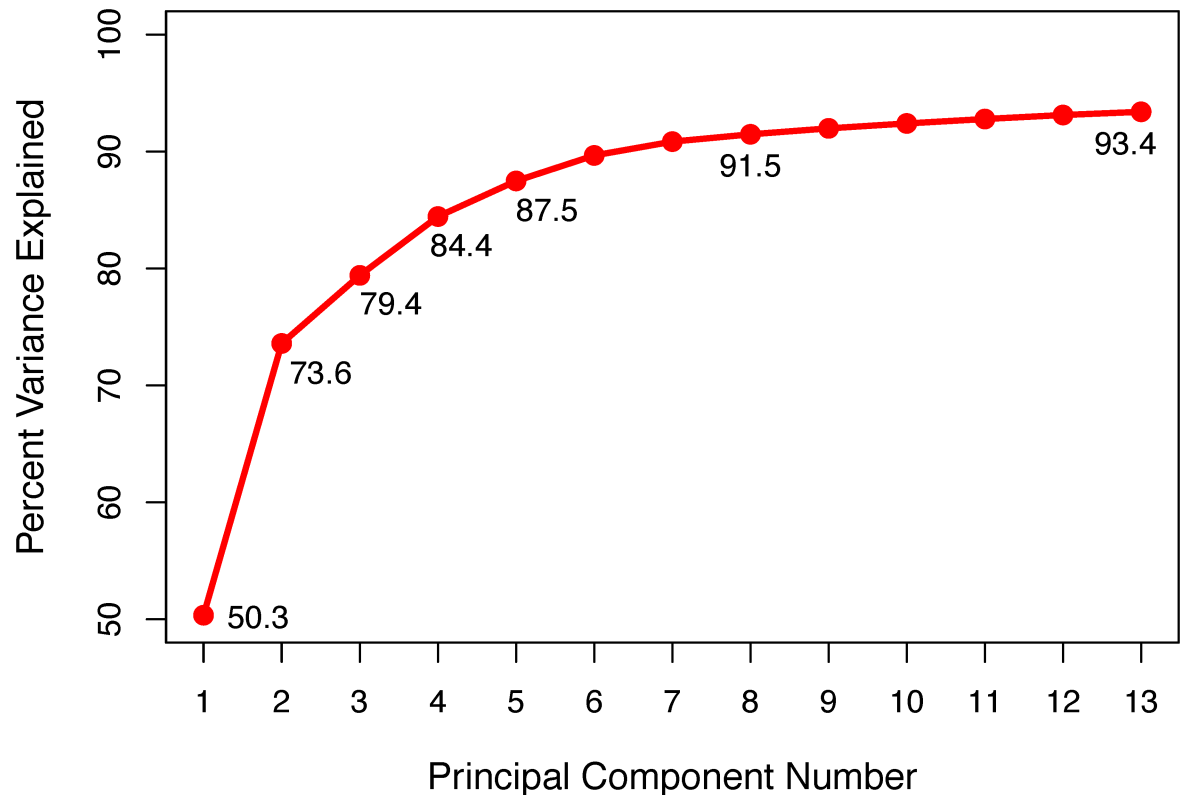


Fig 7. Percent genetic variance explained by principal component analysis. The percent of genetic variance explained is on the y-axis and the principle component (PC) number is on the x-axis. Exact values of percent variance explained are included next to the plotted points at PCs 1-5, 8, and 13.

<https://doi.org/10.1371/journal.pone.0189277.g007>

sixth candidate region, LD did not decay to the same level until well after 10 kbp. In another study that examined sequence variants (SNPs and indels) across 18 gene regions in 36 elite U.S. maize inbreds, Ching et al., (2002) [70] concluded that linkage disequilibrium does not significantly decay within the analyzed range of 300-500 bp. Palaisa et al., (2003) [71] found that LD surrounding two loci in a group of 82 diverse inbred lines decayed to a level of r^2 at 0.2 at approximately 1,000 bp. Similarly, in a study of the *adh1* locus within 32 elite North American public and proprietary inbred lines, Jung et al., (2004) [72] reported that while measurable levels of LD appeared to persist past 500 kbp, it could not be stated whether these long-range regions of sustained LD are common. (Estimates of LD at the specific level of r^2 at 0.2 were not available in this study.)

More recently, Yan et al. (2009) [73] studied the extent of LD in 632 maize inbreds from temperate, tropical, and subtropical regions, using 1,229 SNP markers. They found that across all 632 inbreds, LD decayed to an average r^2 of 0.2 at about 500 bp, and generally concluded that the distance of LD decay is much higher in temperate inbreds than in tropical or subtropical inbreds. Truntzler et al., (2012) [74] using a mix of inbreds from public institutions (113 inbreds) and private companies (201 inbreds) and 979 polymorphic SNP markers, reported that while there is a faster rate of LD decay for the private inbreds than the public inbreds, both sets reach r^2 of 0.2 at a distance of about 1-3 kbp. Romay et al. (2013) [38], using a population that is essentially a subset of the population used in this study, found that while the LD for 212 ex-PVP inbreds declined to an average r^2 of 0.2 at 10 kb, the LD decay among public inbreds was much more rapid, reaching r^2 of 0.2 at 1 kb.

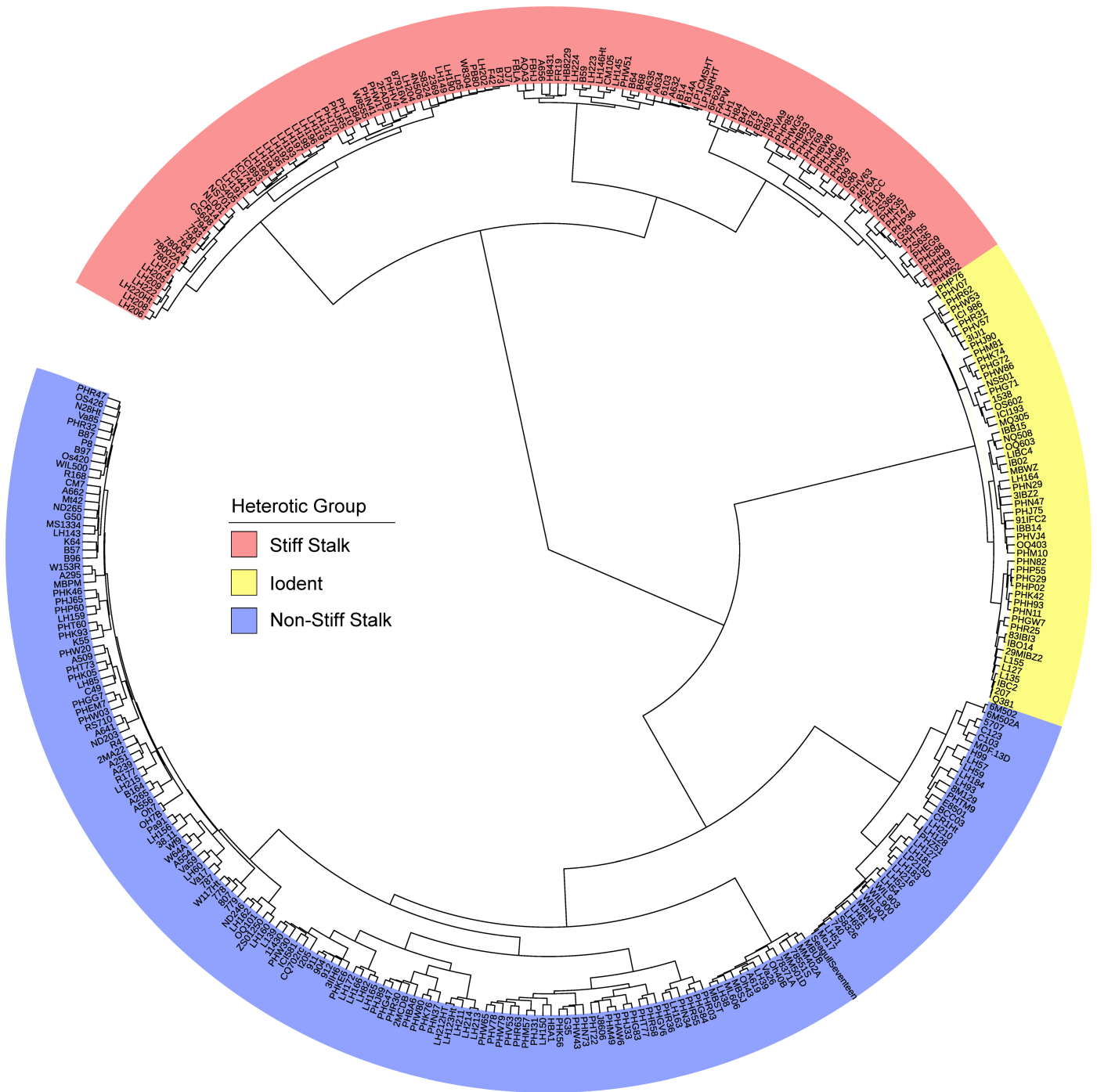


Fig 8. Dendrogram of ex-PVP and public founder inbreds. Circular dendrogram of ex-PVP and public founder inbreds, divided into three heterotic groups. This dendrogram, shown with relative scaled branch lengths and colored according to generally known maize heterotic groups, is based on a cluster analysis using Ward's minimum distance variance method on the matrix of Nei's genetic distance [57, 58]. Scaled branch lengths allow a visual representation of the relative proportion of genetic difference between the three main heterotic groups. Consultation of available pedigrees confirm the accuracy of heterotic group placement for individual inbreds [12, 20, 24, 48, 49, 67]. Note: this tree is presented in a rooted format with the primary purpose of illustrating genetic distance while retaining legible inbred names. While no inference is made about common ancestors, the Stiff Stalk and Iodent/Non-Stiff Stalk portions form an ingroup/outgroup interaction, thus ensuring that the presentation of a tree in rooted format is still an acceptable depiction of the detailed population stratification.

<https://doi.org/10.1371/journal.pone.0189277.g008>

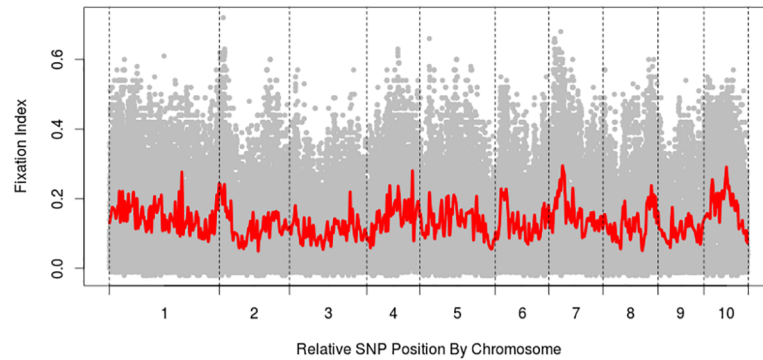


Fig 9. Fixation index values across the genome. Relative SNP position by chromosome (x-axis) and fixation index (y-axis). Each grey dot represents the fixation index statistic (F_{ST}) for an individual genetic locus. High F_{ST} value for a genetic locus may indicate that particular genetic locus contributed to genetic differentiation between heterotic groups. The red trend line represents a moving average across a window of 70 SNPs, or approximately 3570 Mb. This red line is representative, then, of the F_{ST} values across genomic regions. Peaks observed in the trend line, particularly in chromosomes 1, 4, 7, and 10, may be indicative of lengthy genomic regions contributing to heterosis observed in hybrid crosses between inbreds from different heterotic groups.

<https://doi.org/10.1371/journal.pone.0189277.g009>

The calculation of decay of linkage disequilibrium is affected by many factors: composition of the germplasm set; marker characteristics such as quality and genome coverage; and the analysis method employed. While these factors may lead to variance between data sets of the physical distance observed at the standard-reported linkage equilibrium value r^2 of 0.2, the general trends found in this germplasm set are consistent with the results previously reported.

Population structure

Inclusion of a scale in the dendrogram (Fig 8) allows inferences to be made about relative genetic distance between heterotic groups. A comparison between the three heterotic groups

Table 5. Candidate genes for SNPs with high F_{ST} values.

Chr.	Position ^a	F_{ST}	Gene ID	Gene Product Description
2	4209922	0.6216	AC210003.2_FG004	Peroxidase 16
			GRMZM2G045049	FGAM synthase
			GRMZM2G342628	-
2	7744913	0.7442	GRMZM2G180870	Glycosyl hydrolases fam. 16 protein
2	11060537	0.6396	GRMZM2G070468	Transferase
			GRMZM2G111232	Protein phosphatase 2C
4	166831567	0.6396	GRMZM2G003501	G14a
5	32460265	0.6686	GRMZM2G139024	Transcription factor Dp-1
7	17142974	0.6657	GRMZM2G019443	AP2 domain cont. protein RAP2.11
			27658449	0.6537
			GRMZM2G162267	-
7	50367063	0.6877	GRMZM2G072868	receptor kinase
7	92941759	0.6023	GRMZM2G129420	Mitochon. ATP synthase subunit
			GRMZM2G129453	Desaturase/cytochrome b5 protein
8	161042132	0.6033	GRMZM2G330719	Cons. gene of unknown function
8	172812044	0.6033	GRMZM2G076410	-

^aPhysical position of the marker within the specified chromosome, in bp.

<https://doi.org/10.1371/journal.pone.0189277.t005>

Table 6. Genomic regions with high F_{ST} values.

Chr.	First Marker	Last Marker	Region Size	Mean F_{ST}
1	242575005	245035347	2460342	0.2787
4	166907041	168546722	1639681	0.2885
7	36258886	41462768	5203882	0.3039
10	116450442	117756633	1306191	0.2991

First Marker, Last Marker, and Region Size are all given in bp.

<https://doi.org/10.1371/journal.pone.0189277.t006>

in Fig 8 reinforces the concept that commercial breeding efforts in the Plant Variety Protection era (post-1970) have continued to drive genetic divergence among Stiff Stalk, Non-Stiff Stalk, and Iodent. F_{ST} analysis confirms the divergence, with the result of 0.1732 indicating moderate levels of divergence among these three heterotic groups. A more detailed look at the F_{ST} analysis reveals several genomic regions as well as several individual SNPs with high F_{ST} values. High F_{ST} values for a particular region or SNP mean that it is more likely that the major haplotype or allele in the SS group is different than the major haplotype or allele in the NSS group. These genomic regions and individual SNPs that show high genetic diversity between heterotic groups deserve further study, as they may provide insights about the genetic basis of heterosis.

It is widely accepted among breeders and others familiar with North American maize germplasm that heterotic groups continue to diverge genetically [36, 75]. One reason for this genetic divergence of heterotic groups could be the widespread breeding practice of recycling of elite inbreds within heterotic groups to produce new inbreds, then evaluating them based on test-cross performance with inbreds from other heterotic groups. The observed genetic divergence between breeding pools then may be a response to selection for heterosis in testcross hybrids. The value of a commercial inbred is not just based on its ability to efficiently produce hybrid seed, but primarily on its ability to consistently produce superior grain yield in a testcross. Therefore, as inbreds are judged by their performance in a testcross, a higher degree of genetic divergence between heterotic groups may be a result of selection over time for better hybrid performance.

Application of genetic relationships in breeding

Precise and accurate knowledge of the genetic background of a particular inbred can be very useful to a plant breeder in determining the best use of that inbred. Traditional pedigree information, supplemented by population genetics data can help a breeder decide what combination of inbreds may prove to be the best for breeding crosses and for hybrid testcrosses. Many PVP inbreds came from self-pollination of commercial hybrids. An accurate dendrogram based on genetic relationships can help breeders better understand the genetic background of PVP inbreds derived from commercial hybrids, as well as identify close genetic relatives. For one example, the P3737-derived inbreds 3IIH6, 912, 904, and 911 are located near the bottom of the tree in S9 Fig. The location of these lines within the dendrogram does not align with expectations based only on field testing and general pedigree knowledge. With robust genetic diversity analysis, however, a more clear and complete picture emerges.

The dendrogram produced in this study (Fig 8) visually identifies the heterotic group membership of each ex-PVP and public inbred. The divisions among Stiff Stalk, Non-Stiff Stalk, and Iodent are clear. Further sub-group divisions within the Stiff Stalk heterotic group are defined. However, the sub-group divisions within the Non-Stiff Stalk heterotic group are more difficult to resolve (see S1 Fig). Many of the ex-PVP Non-Stiff Stalk inbreds are genetically closely

related, especially in the “Pioneer Mixed” and “Miscellaneous” groups. This study includes a larger number of ex-PVP inbreds and more detailed information about relationships derived directly from genotypic cluster analysis than previous studies. In general, the results presented here agree with previous classifications of maize heterotic groups [21, 25, 36, 47].

Information from this study can be useful in determining how to begin testing a newly released ex-PVP inbred line. When the PVP certificate for an inbred expires and the seed is freely available for use, the parentage of the line can be determined by consulting the pedigree on the certificate. Then the parental inbreds can be located on the dendrogram. Thus, the newly released ex-PVP inbred can be anchored to previously characterized inbreds. Such an approach can potentially save time and resources, particularly for smaller breeding programs.

Previous yield trial results of parental lines could be a logical starting point for determining the potential combining patterns and agronomic performance value of a newly released ex-PVP inbred [76]. Alternatively, if the pedigree on the certificate does not include parental inbreds that are within the current genetic cluster diagram, and if the inbred can be quickly genotyped, then the inbred can be included in a new cluster analysis where the precise genetic relationships can be determined. Even if good parental pedigree and testcross data is available for a newly expired PVP inbred, there may be merit to genotyping the inbred and determining where it falls in the cluster diagram, as this provides complementary and more precise genetic relationship information. Yield trial data coupled with this population genetic analysis may further improve a breeder’s ability to immediately identify the best material and quickly integrate it into a germplasm pool. Understanding the genetic relationships and population differentiation of elite maize germplasm is an integral part of helping breeders to maintain and potentially increase the rate of genetic gain, resulting in higher overall agronomic performance of inbreds and hybrids.

Supporting information

S1 Table. Ex-PVP inbreds used in this study.

(CSV)

S2 Table. Public inbreds used in this study.

(CSV)

S3 Table. Subsets of inbreds used in F_{ST} calculations.

(CSV)

S4 Table. List of individual accession identifiers for all inbreds used in this study for genotyping-by-sequencing (GBS) data, available at www.panzea.org [50]. In the data repository, individual accession identifiers are referred to as “Taxa”, and thus have been listed as such in this table.

(CSV)

S1 Fig. Dendrogram of ex-PVP and public founder inbreds, divided into eight heterotic groups. Shown with relative scaled branch lengths, this dendrogram is based on a cluster analysis using Ward’s minimum distance variance method, and Nei’s genetic distance [57, 58].

Colors represent further divisions of heterotic groups of maize, with groups named by important founder line or by general group composition. Consultation of published pedigrees [48, 49, 67] as well as previous publications on the subject as well as previous publications on the subject [12, 20, 24] confirm the accuracy of heterotic group placement for individual inbreds.

(TIFF)

S2 Fig. Linear dendrogram of ex-PVP and public founder inbreds, divided into eight subgroups. This dendrogram is based on phylogenetic cluster analysis using Ward's minimum distance variance method, and Nei's genetic distance [57, 58]. Tree branch lengths are scaled relatively according to the actual genetic distance matrix. Colors correlate with maize family groups as indicated in the "Heterotic Group" key. Pedigrees are included to the right of each inbred. PVP inbred pedigrees were obtained from from PVP certificates, available at ars.grin.gov [48]. Public inbred pedigrees were obtained from Gerdes et al., (1993) [49] and Cross et al., (1989) [67]. Consultation of pedigrees, as well as previous publications on the subject [12, 20, 24], confirm individual heterotic group memberships are accurate.
(TIFF)

S3 Fig. B73 group dendrogram with inbred pedigrees. The color surrounding the ex-PVP and public inbred names corresponds with the color assigned to each family subgroup in S1 and S2 Figs. Pedigrees are included to the right of each inbred. PVP inbred pedigrees were obtained from from PVP certificates, available at ars.grin.gov [48]. Public inbred pedigrees were obtained from Gerdes et al., (1993) [49] and Cross et al., (1989) [67]. Consultation of pedigrees, as well as previous publications on the subject [12, 20, 24], confirm individual heterotic group memberships are accurate.
(TIFF)

S4 Fig. B14 group dendrogram with inbred pedigrees. The color surrounding the ex-PVP and public inbred names corresponds with the color assigned to each family subgroup in S1 and S2 Figs. Pedigrees are included to the right of each inbred. PVP inbred pedigrees were obtained from from PVP certificates, available at ars.grin.gov [48]. Public inbred pedigrees were obtained from Gerdes et al., (1993) [49] and Cross et al., (1989) [67]. Consultation of pedigrees, as well as previous publications on the subject [12, 20, 24], confirm individual heterotic group memberships are accurate.
(TIFF)

S5 Fig. B37 group dendrogram with inbred pedigrees. The color surrounding the ex-PVP and public inbred names corresponds with the color assigned to each family subgroup in S1 and S2 Figs. Pedigrees are included to the right of each inbred. PVP inbred pedigrees were obtained from from PVP certificates, available at ars.grin.gov [48]. Public inbred pedigrees were obtained from Gerdes et al., (1993) [49] and Cross et al., (1989) [67]. Consultation of pedigrees, as well as previous publications on the subject [12, 20, 24], confirm individual heterotic group memberships are accurate.
(TIFF)

S6 Fig. Iodent group dendrogram with inbred pedigrees. The color surrounding the ex-PVP and public inbred names corresponds with the color assigned to each family subgroup in S1 and S2 Figs. Pedigrees are included to the right of each inbred. PVP inbred pedigrees were obtained from from PVP certificates, available at ars.grin.gov [48]. Public inbred pedigrees were obtained from Gerdes et al., (1993) [49] and Cross et al., (1989) [67]. Consultation of pedigrees, as well as previous publications on the subject [12, 20, 24], confirm individual heterotic group memberships are accurate.
(TIFF)

S7 Fig. Lancaster group dendrogram with inbred pedigrees. The color surrounding the ex-PVP and public inbred names corresponds with the color assigned to each family subgroup in S1 and S2 Figs. Pedigrees are included to the right of each inbred. PVP inbred pedigrees were obtained from from PVP certificates, available at ars.grin.gov [48]. Public inbred pedigrees

were obtained from Gerdes et al., (1993) [49] and Cross et al., (1989) [67]. Consultation of pedigrees, as well as previous publications on the subject [12, 20, 24], confirm individual heterotic group memberships are accurate.

(TIFF)

S8 Fig. Ohio 43 group dendrogram with inbred pedigrees. The color surrounding the ex-PVP and public inbred names corresponds with the color assigned to each family subgroup in S1 and S2 Figs. Pedigrees are included to the right of each inbred. PVP inbred pedigrees were obtained from from PVP certificates, available at ars.grin.gov [48]. Public inbred pedigrees were obtained from Gerdes et al., (1993) [49] and Cross et al., (1989) [67]. Consultation of pedigrees, as well as previous publications on the subject [12, 20, 24], confirm individual heterotic group memberships are accurate.

(TIFF)

S9 Fig. Pioneer Mixed group dendrogram with inbred pedigrees. The color surrounding the ex-PVP and public inbred names corresponds with the color assigned to each family subgroup in S1 and S2 Figs. Pedigrees are included to the right of each inbred. PVP inbred pedigrees were obtained from from PVP certificates, available at ars.grin.gov [48]. Public inbred pedigrees were obtained from Gerdes et al., (1993) [49] and Cross et al., (1989) [67]. Consultation of pedigrees, as well as previous publications on the subject [12, 20, 24], confirm individual heterotic group memberships are accurate.

(TIFF)

S10 Fig. Miscellaneous sub-group dendrogram with pedigrees. The color surrounding the ex-PVP and public inbred names corresponds with the color assigned to each family subgroup in S1 and S2 Figs. Pedigrees are included to the right of each inbred. PVP inbred pedigrees were obtained from from PVP certificates, available at ars.grin.gov [48]. Public inbred pedigrees were obtained from Gerdes et al., (1993) [49] and Cross et al., (1989) [67]. Consultation of pedigrees, as well as previous publications on the subject [12, 20, 24], confirm individual heterotic group memberships are accurate.

(TIFF)

Acknowledgments

The authors gratefully acknowledge the help of Alencar Xavier and William Muir, coauthors of the R software package ‘NAM’, who provided assistance in overall coding and analysis, and guidance with interpretation of F_{ST} statistics.

Author Contributions

Conceptualization: Travis J. Beckett, A. Jason Morales, Klaus L. Koehler.

Data curation: Travis J. Beckett, A. Jason Morales, Klaus L. Koehler.

Formal analysis: Travis J. Beckett, A. Jason Morales.

Funding acquisition: Travis J. Beckett, A. Jason Morales, Klaus L. Koehler, Torbert R. Rocheford.

Investigation: Travis J. Beckett, A. Jason Morales.

Methodology: Travis J. Beckett, A. Jason Morales, Klaus L. Koehler.

Project administration: Travis J. Beckett, A. Jason Morales, Klaus L. Koehler.

Resources: Travis J. Beckett, A. Jason Morales, Klaus L. Koehler.

Software: Travis J. Beckett, A. Jason Morales.

Supervision: Travis J. Beckett, Klaus L. Koehler, Torbert R. Rocheford.

Validation: Travis J. Beckett, A. Jason Morales.

Visualization: Travis J. Beckett, A. Jason Morales.

Writing – original draft: Travis J. Beckett.

Writing – review & editing: Travis J. Beckett, A. Jason Morales, Klaus L. Koehler, Torbert R. Rocheford.

References

1. Shull GH. A pure-line method in corn breeding. *Journal of Heredity*. 1909 Jan 1(1):51–8. <https://doi.org/10.1093/jhered/os-5.1.51>
2. Shull GH. Hybridization methods in corn breeding. *Journal of Heredity*. 1910 Apr 1; 1(2):98–107. <https://doi.org/10.1093/jhered/1.2.98>
3. Richey FD. The experimental basis for the present status of corn breeding. *Agronomy Journal*. 1922; 14(1-2):1–7. <https://doi.org/10.2134/agronj1922.00021962001401-20001x>
4. East EM, Hayes HK. Heterozygosis in evolution and in plant breeding. US Government Printing Office; 1912.
5. East EM. Inheritance in maize. *Botanical Gazette*. 1913 May 1; 55(5):404–5. <https://doi.org/10.1086/331078>
6. Hayes HK, East EM. Improvement in corn. Connecticut Agricultural Experiment Station; 1911.
7. Hayes HK, East EM. Further experiments on inheritance in maize. Connecticut Agricultural Experiment Station; 1915.
8. Wallace HA, Brown WL. Corn, And Its Early Fathers. *Soil Science*. 1957 Feb 1; 83(2):160. <https://doi.org/10.1097/00010694-195702000-00014>
9. Crabb AR. The hybrid-corn makers. Prophets of plenty. 1947.
10. USDA United States Department of Agriculture, National Agricultural Statistics Service: Quick Stats [Online Database]. National Agricultural Statistics Service, Washington D.C.; 2016.
11. USDA. United States Department of Agriculture: Plant Variety Protection Act and Regulations and Rules of Practice. Agricultural Marketing Service, Washington, DC, updated July 2013.
12. Mikel MA. Availability and analysis of proprietary dent corn inbred lines with expired US plant variety protection. *Crop Science*. 2006 Nov 1; 46(6):2555–60.
13. *Diamond v. Chakrabarty*, 447 US 303 (1980).
14. *J.E.M. AG Supply v. Pioneer Hi-Bred International*, 122 US 593 (2001).
15. United States Patent and Trademark Office. Patentable Subject Matter—Living Subject Matter. Manual of Patent Examining Procedure, ch. 2100 sec. 2105. Jul 2015. URL <https://www.uspto.gov/web/offices/pac/mpep/s2105.html>.
16. *Ex parte Hibberd*, 227 USPQ 433. (Bd. Pat. App. & Inter. 1985).
17. Van Brunt J. *Ex parte Hibberd*: Another Landmark Decision. *Nature Biotechnology*. 1985 Dec 1; 3(12):1059–60. <https://doi.org/10.1038/nbt1285-1059>
18. Moschini GC. Competition issues in the seed industry and the role of intellectual property. [www.ChoicesMagazine.org](http://www.choicesmagazine.org/UserFiles/file/article_120.pdf). 2nd Quarter 2010; 25(2). URL http://www.choicesmagazine.org/UserFiles/file/article_120.pdf.
19. Kurtz B, Gardner CA, Millard MJ, Nickson T, Smith JS. Global Access to Maize Germplasm Provided by the US National Plant Germplasm System and by US Plant Breeders. *Crop Science*. 2016; 56(3):931–41. <https://doi.org/10.2135/cropsci2015.07.0439>
20. Mikel MA. Genetic composition of contemporary US commercial dent corn germplasm. *Crop Science*. 2011 Mar 1; 51(2):592–9. <https://doi.org/10.2135/cropsci2010.06.0332>
21. Bernardo R. Essentials of plant breeding. 2014.
22. Troyer AF. Background of U.S. hybrid corn. *Crop Science*. 1999; 39(3):601–26. <https://doi.org/10.2135/cropsci1999.0011183X003900020001x>

23. Gethi JG, Labate JA, Lamkey KR, Smith ME, Kresovich S. SSR variation in important US maize inbred lines. *Crop Science*. 2002 May 1; 42(3):951–7. <https://doi.org/10.2135/cropsci2002.0951>
24. Mikel MA, Dudley JW. Evolution of North American dent corn from public to proprietary germplasm. *Crop Science*. 2006 May 1; 46(3):1193–205. <https://doi.org/10.2135/cropsci2005.10-0371>
25. Nelson PT, Coles ND, Holland JB, Bubeck DM, Smith S, Goodman MM. Molecular characterization of maize inbreds with expired US plant variety protection. *Crop Science*. 2008 Sep 1; 48(5):1673–85. <https://doi.org/10.2135/cropsci2008.02.0092>
26. Lu Y, Yan J, Guimaraes CT, Taba S, Hao Z, Gao S, et al. Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. *Theoretical and Applied Genetics*. 2009 Dec 1; 120(1):93–115. <https://doi.org/10.1007/s00122-009-1162-7> PMID: 19823800
27. Fan XM, Zhang YM, Yao WH, Chen HM, Tan J, Xu CX, et al. Classifying maize inbred lines into heterotic groups using a factorial mating design. *Agronomy Journal*. 2009 Jan 1; 101(1):106–12. <https://doi.org/10.2134/agronj2008.0217>
28. Smith OS, Smith JS, Bowen SL, Tenborg RA, Wall SJ. Similarities among a group of elite maize inbreds as measured by pedigree, F1 grain yield, grain yield, heterosis, and RFLPs. *Theoretical and Applied Genetics*. 1990 Dec 1; 80(6):833–40. <https://doi.org/10.1007/BF00224201> PMID: 24221118
29. Dudley JW, Maroof MA, Rufener GK. Molecular markers and grouping of parents in maize breeding programs. *Crop Science*. 1991; 31(3):718–23. <https://doi.org/10.2135/cropsci1991.0011183X003100030036x>
30. Livini C, Ajmone-Marsan P, Melchinger AE, Messmer MM, Motto M. Genetic diversity of maize inbred lines within and among heterotic groups revealed by RFLPs. *Theoretical and Applied Genetics*. 1992 Jun 1; 84(1-2):17–25. <https://doi.org/10.1007/BF00223976> PMID: 24203023
31. Melchinger AE, Messmer MM, Lee M, Woodman WL, Lamkey KR. Diversity and relationships among US maize inbreds revealed by restriction fragment length polymorphisms. *Crop Science*. 1991; 31(3):669–78. <https://doi.org/10.2135/cropsci1991.0011183X003100030025x>
32. Mumm RH, Dudley JW. A classification of 148 US maize inbreds: I. Cluster analysis based on RFLPs. *Crop Science*. 1994; 34(4):842–51. <https://doi.org/10.2135/cropsci1994.0011183X003400040006x>
33. Senior ML, Murphy JP, Goodman MM, Stuber CW. Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. *Crop Science*. 1998; 38(4):1088–98. <https://doi.org/10.2135/cropsci1998.0011183X003800040034x>
34. Barata C, Carena MJ. Classification of North Dakota maize inbred lines into heterotic groups based on molecular and testcross data. *Euphytica*. 2006 Oct 1; 151(3):339–49. <https://doi.org/10.1007/s10681-006-9155-y>
35. Kahler AL, Kahler JL, Thompson SA, Ferriss RS, Jones ES, Nelson BK, et al. North American study on essential derivation in maize: II. Selection and evaluation of a panel of simple sequence repeat loci. *Crop Science*. 2010 Mar 1; 50(2):486–503. <https://doi.org/10.2135/cropsci2009.03.0121>
36. van Heerwaarden J, Hufford MB, Ross-Ibarra J. Historical genomics of North American maize. *Proceedings of the National Academy of Sciences*. 2012 Jul 31; 109(31):12420–5. <https://doi.org/10.1073/pnas.1209275109>
37. Olmos SE, Delucchi C, Ravera M, Negri ME, Mandolino C, Eyherabide GH. Genetic relatedness and population structure within the public Argentinean collection of maize inbred lines. *Maydica*. 2013 Oct 29; 59(1):16–31.
38. Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM, et al. Comprehensive genotyping of the USA national maize inbred seed bank. *Genome biology*. 2013 Jun 11; 14(6):1. <https://doi.org/10.1186/gb-2013-14-6-r55>
39. Unterseer S, Bauer E, Haberer G, Seidel M, Knaak C, Ouzunova M, et al. A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array. *BMC Genomics*. 2014 Sep 29; 15(1):1. <https://doi.org/10.1186/1471-2164-15-823>
40. Smith SD, Murray SC, Heffner E. Molecular analysis of genetic diversity in a Texas maize (*Zea mays L.*) breeding program. *Maydica*. 2015 Jul 17; 60(2015):M20.
41. Wu Y, San Vicente F, Huang K, Dhliwayo T, Costich DE, Semagn K, et al. Molecular characterization of CIMMYT maize inbred lines with genotyping-by-sequencing SNPs. *Theoretical and Applied Genetics*. 2016 Apr 1; 129(4):753–65. <https://doi.org/10.1007/s00122-016-2664-8> PMID: 26849239
42. Zhang X, Zhang H, Li L, Lan H, Ren Z, Liu D, et al. Characterizing the population structure and genetic diversity of maize breeding germplasm in Southwest China using genome-wide SNP markers. *BMC genomics*. 2016 Aug 31; 17(1):697. <https://doi.org/10.1186/s12864-016-3041-3> PMID: 27581193
43. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one*. 2011 May 4; 6(5):e19379. <https://doi.org/10.1371/journal.pone.0019379> PMID: 21573248

44. Deschamps S, Llaca V, May GD. Genotyping-by-sequencing in plants. *Biology*. 2012 Sep 25; 1(3):460–83. <https://doi.org/10.3390/biology1030460> PMID: 24832503
45. Poland JA, Rife TW. Genotyping-by-sequencing for plant breeding and genetics. *The Plant Genome*. 2012 Nov 1; 5(3):92–102.
46. Swarts K, Li H, Romero Navarro JA, An D, Romay MC, Hearne S, et al. Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. *The Plant Genome*. 2014 Nov 1; 7(3). <https://doi.org/10.3835/plantgenome2014.05.0023>
47. Morales AJ. Genomic approaches for improving grain yield in maize using formerly plant variety protected germplasm (Doctoral dissertation, PURDUE UNIVERSITY), 2013.
48. USDA Germplasm Resources Information Network-(GRIN)[Online Database]. USDA, ARS, National Genetic Resources Program, National Germplasm Resources Laboratory, Beltsville, Maryland, 2013.
49. Gerdes JT, Behr CF, Coors HG, Tracy WF, Avratovscukova N. Compilation of North American maize breeding germplasm. Madison: CSSA; 1993.
50. Zhao W, Canaran P, Jurkuta R, Fulton T, Glaubitz J, Buckler E, et al. Panzea: a database and resource for molecular and functional diversity in the maize genome. *Nucleic Acids Research*. 2006 Jan 1; 34 (suppl 1):D752–7. <https://doi.org/10.1093/nar/gkj011> PMID: 16381974
51. Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, et al. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS One*. 2014 Feb 28; 9(2):e90346. <https://doi.org/10.1371/journal.pone.0090346> PMID: 24587335
52. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. 2007 Oct 1; 23(19):2633–5. <https://doi.org/10.1093/bioinformatics/btm308> PMID: 17586829
53. He S, Zhao Y, Mette MF, Bothe R, Ebmeyer E, Sharbel TF, et al. Prospects and limits of marker imputation in quantitative genetic studies in European elite wheat (*Triticum aestivum* L.). *BMC Genomics*. 2015 Mar 11; 16(1):1. <https://doi.org/10.1186/s12864-015-1366-y>
54. Xavier A, Xu S, Muir WM, Rainey KM. NAM: association studies in multiple populations. *Bioinformatics*. 2015 Dec 1; 31(23):3862–4. PMID: 26243017
55. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2015.
56. Herrmann E, Mächler M. lokern: Kernel Regression Smoothing with Local or Global Plug-in Bandwidth, 2013. URL <http://CRAN.R-project.org/package=lokern>. R package version 2013:1-1.
57. Nei M. Genetic distance between populations. *American Naturalist*. 1972 May 1:283–92. <https://doi.org/10.1086/282771>
58. Ward JH Jr. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*. 1963 Mar 1; 58(301):236–44. <https://doi.org/10.1080/01621459.1963.10500845>
59. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*. 2004 Jan 22; 20(2):289–90. <https://doi.org/10.1093/bioinformatics/btg412> PMID: 14734327
60. Letunic I, Bork P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research*. 2011 Apr 5:gkr201. <https://doi.org/10.1093/nar/gkr201>
61. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution*. 1984 Nov 1:1358–70. <https://doi.org/10.2307/2408641> PMID: 28563791
62. Porto-Neto LR, Lee SH, Lee HK, Gondro C. Detection of signatures of selection using F ST. *Genome-Wide Association Studies and Genomic Prediction*. 2013:423–36. https://doi.org/10.1007/978-1-62703-447-0_19
63. Ziegler GR, Hartssock RH, Baxter I. Zbrowse: an interactive GWAS results browser. *PeerJ Computer Science*. 2015 May 27; 1:e3. <https://doi.org/10.7717/peerj-cs.3>
64. Thorndike RL. Who belongs in the family? *Psychometrika*. 1953 Dec 27; 18(4):267–76. <https://doi.org/10.1007/BF02289263>
65. Kaufman L, Rousseeuw PJ. Finding groups in data: an introduction to cluster analysis. John Wiley & Sons; 2009 Sep 25.
66. Hastie T, Tibshirani R, Walther G. Estimating the number of data clusters via the Gap statistic. *J Roy Stat Soc B*. 2001 Aug; 63:411–23.
67. Cross HZ. ND265: a new parental line of early corn. *North Dakota Farm Research*. 1989; 47(3):19–21.
68. Tenailon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* subsp. *mays* L.). *Proceedings of the National Academy of Sciences*. 2001 Jul 31; 98(16):9161–6. <https://doi.org/10.1073/pnas.151244298>

69. Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, et al. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences*. 2001 Sep 25; 98(20):11479–84. <https://doi.org/10.1073/pnas.201394398>
70. Ching AD, Caldwell KS, Jung M, Dolan M, Smith O, Tingey S, et al. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC genetics*. 2002 Oct 7; 3(1):19. <https://doi.org/10.1186/1471-2156-3-19> PMID: 12366868
71. Palaisa KA, Morgante M, Williams M, Rafalski A. Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. *The Plant Cell*. 2003 Aug 1; 15(8):1795–806. <https://doi.org/10.1105/tpc.012526> PMID: 12897253
72. Jung M, Ching A, Bhatramakki D, Dolan M, Tingey S, Morgante M, et al. Linkage disequilibrium and sequence diversity in a 500-kbp region around the *adh1* locus in elite maize germplasm. *Theoretical and Applied Genetics*. 2004 Aug 1; 109(4):681–9. <https://doi.org/10.1007/s00122-004-1695-8> PMID: 15300382
73. Yan J, Shah T, Warburton ML, Buckler ES, McMullen MD, Crouch J. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PloS one*. 2009 Dec 24; 4(12):e8451. <https://doi.org/10.1371/journal.pone.0008451> PMID: 20041112
74. Truntzler M, Ranc N, Sawkins MC, Nicolas S, Manicacci D, Lespinasse D, et al. Diversity and linkage disequilibrium features in a composite public/private dent maize panel: consequences for association genetics as evaluated from a case study using flowering time. *Theoretical and Applied Genetics*. 2012 Aug 1; 125(4):731–47. <https://doi.org/10.1007/s00122-012-1866-y> PMID: 22622520
75. Lamkey CM, Lorenz AJ. Relative effect of drift and selection in diverging populations within a reciprocal recurrent selection program. *Crop Science*. 2014; 54(2):576–85. <https://doi.org/10.2135/cropsci2013.07.0484>
76. Beckett TJ. Analysis of Genetic Loci Associated with Agronomic Performance in Previously Plant-Variety-Protected Elite Commercial Maize Germplasm (Master's thesis), PURDUE UNIVERSITY, 2016.