Genetic diversity of North American popcorn germplasm and the effect of population structure on nicosulfuron response

Madsen Sullivan[1], Martin M. Williams II[2], Anthony J. Studer[1,†]

Affiliations:

[1] Department of Crop Sciences, University of Illinois, Urbana, IL 61801, USA.

[2] Global Change and Photosynthesis Research Unit, USDA-ARS, Urbana, IL 61801, USA.

[†] Corresponding author (astuder@illinois.edu).

Abbreviations:

ALS, acetolactate synthase; BCAA, branched-chain amino acid; CIMMYT, International Maize and Wheat Improvement Center; DAPC, discriminant analysis of principal components; GBS, genotyping-by-sequencing; GRIN, Germplasm Resources Information Network; GWAS, genome-wide association studies; IBS, identity by state; LD, linkage disequilibrium; MAF, minor allele frequency; MAPQ, mapping quality; PCoA, principal coordinate analysis; SNP, single nucleotide polymorphism

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/csc2.21039.

ABSTRACT

Popcorn (*Zea mays*) is an important crop in the United States; however, genetic analyses of popcorn are limited and tend to utilize relatively few markers that cannot capture the total genomic variation. To improve the genomic resources in popcorn, a panel of 320 popcorn accessions was evaluated using 308,811 single nucleotide polymorphisms generated using a genotyping-by-sequencing approach. Using this genomic data, several model-based clustering analyses identified two major groups. The first comprised North American Yellow Pearl Popcorns and accessions of the Chilean Curagua landrace, separated into three subgroups. The second, the Pointed and Latin American Popcorns, included all remaining North American, Latin American, and global accessions. These groups exhibited differences in population structure and genetic diversity. The North American Yellow Pearl Popcorns contain limited genetic diversity compared to the Pointed and Latin American Popcorns. Additionally, phenotypic differences between the two groups were observed in kernel color and nicosulfuron sensitivity. A filtered set of SNPs was curated and used for genome-wide association studies and popcorn-specific candidate genes for nicosulfuron tolerance were identified. The genomic characterization described here can be used by breeding programs to accelerate the rate of genetic gain and incorporate genetic diversity into elite popcorn germplasm.

1 INTRODUCTION

Popcorn (*Zea mays*) is a popular snack and a valuable commercial crop in the United States. In a global market worth 10 billion USD, the United States has the highest per capita popcorn consumption of any country at over 55 liters per year, representing a total of 16 billion liters consumed annually (Serna-Saldivar 2022). Likewise, the United States grows more popcorn than any

other nation, with approximately 90,000 hectares of popcorn planted annually, most of which are grown in Nebraska, Indiana, Ohio, and Illinois (USDA NASS, 2022).

Popcorn has been cultivated within the territory of the modern United States since pre-Columbian times, although the range of pre-Columbian popcorn was limited to the American Southwest. Popcorn landraces first reached this region by 4,200 BP (Merrill et al. 2009), likely first from a highland origin, including the landraces Palomero Toluqueño and Palomero de Jalisco, followed by a later mixing with lowland landraces, such as Chapalote and Reventador (da Fonseca et al. 2015; Wang et al. 2017). Although these landraces do pop, the popping quality is poor compared to modern lines in terms of expansion, hull fragmentation, and percentage of unpopped kernels (Vazquez-Carrillo et al. 2019; Bautista-Ramirez et al. 2020).

Despite cultivation in the American Southwest for millennia, popcorn was not widely distributed in North America until the late 18th and early 19th centuries, and even then, the amount grown was limited (Erwin 1950). During the first half of the 19th century, popcorn was typically grown in the United States as a small-scale garden crop for household use (Ziegler 2000). By the late 19th and early 20th century, the popularity of popcorn rose as open-pollinated varieties began reaching more consumers, following the first successful processors and commercial brands in the 1880s and the introduction of the mobile popcorn machine in 1893 (Smith 1999). The predominant types of popcorn grown during this time were the North American Pointed Rice Popcorns or 'hulless' types, characterized as having a pointed upper kernel shape and frequently colored white. The first hybrid popcorn used in commercial production was generated using two selections of Japanese Hulless, a white pointed rice variety (Brunson 1937). However, since the mid-20th century, another group of popcorn has largely replaced the pointed rice type commercially – the North American Yellow Pearl Popcorns. This group has a rounded kernel shape and is predominantly yellow, although color

variation has been introduced by several breeding programs (Ziegler 2000). While the North American Yellow Pearl Popcorns are the most important for commercial production in the United States, North American Pointed Rice Popcorns are still commercially grown and used in breeding programs.

Characterizing the genetic relationships and identifying variation within these distinct groups of North American popcorn is an important aspect of breeding improved popcorn. Early molecular techniques were used to assess the genetic diversity of North American Yellow Pearl Popcorns and identified three heterotic groups – Supergold, South American, and Amber Pearl (Kantety et al. 1995). Later work by Santacruz-Varela et al. defined three major groups of North American popcorn – the aforementioned North American Pointed Rice Popcorn and North American Yellow Pearl Popcorn, as well as the North American Early Popcorn (2004). Additionally, each group's phylogenetic relationship and historical origin were investigated, and possible germplasm sources were identified. The North American Pointed Rice Popcorns appeared to originate from the traditional pointed rice popcorns of Latin America, based on both morphological and genotypic similarities. Furthermore, the North American Yellow Pearl Popcorns were most closely related to, and likely principally derived from, the Chilean Curagua landrace. Finally, the North American Early Popcorns may have been developed from Northern Flint corn and subsequently influenced by European popcorn varieties (Santacruz-Varela et al. 2004). Although this information has been useful in understanding breeding history and relatedness, the limited number of markers and lines has restricted its use in modern breeding programs to heterotic pool development.

Compared with dent corn germplasm, popcorn exhibits considerably inferior agronomic traits, including herbicide tolerance, disease resistance, lodging, yield, and overall vigor (Ziegler 2000). Additionally, differences in herbicide tolerance have been reported within popcorn, with white

popcorn generally exhibiting greater herbicide sensitivity than yellow (Loux et al. 2017). Historically, the poorer agronomic traits of popcorn are due partly to the amount of time and resources breeders have invested relative to dent corn and also because of the emphasis placed on popping quality traits. In addition to agronomic traits, popcorn breeders must consider quality traits directly related to popping and eating quality, including popping expansion, hull separation, frequency of unpopped kernels, texture, and appearance (Matz 1984). Furthermore, dent corn breeding pipelines have integrated molecular characterization and resources, resulting in improved selection and rates of genetic gain, while publicly available resources within popcorn are limited.

High-quality genetic and genomic resources in maize have empowered researchers and facilitated discoveries involving the genetic architecture of traits, candidate gene association and identification, and population structure (Buckler et al. 2009; Cook et al. 2012; Romay et al. 2013; Peiffer et al. 2014; Wallace et al. 2014; Gage et al. 2020). While the genetic resources for mapping populations, diversity panels, and seed banks have been used extensively within dent corn, popcorn germplasm comprises a genetically distinct type of maize that is underrepresented in these studies. Recently, genotyping-by-sequencing (GBS) was performed on popcorn accessions from CIMMYT (Li et al. 2019; Li et al. 2021), Brazilian breeding programs (Senhorinho et al. 2019) and Chinese breeding programs (Yu et al. 2021) and used to characterize genetic diversity, describe population structure, and perform genome-wide association studies (GWAS). The genomic characterization of these diverse popcorn accessions can be used to assist breeders in incorporating genetic variation into breeding programs and identifying loci controlling important traits in popcorn. Likewise, the rate of genetic gain in North American popcorn breeding programs would increase dramatically with the incorporation of high-quality genetic and genomic information, as has been utilized in dent corn breeding for decades.

In this study, molecular analysis was performed on 320 publicly available popcorn accessions genotyped with GBS representing the diversity of North American popcorn. The goals of this study were to i) characterize the genetic diversity of North American popcorn, ii) identify population structure within North American popcorn and other groups of popcorn, iii) describe genomic differences between these groups, iv) perform genome-wide association studies, and v) compare findings with previous attempts to characterize North American popcorn.

2 MATERIALS AND METHODS

2.1 Plant materials and DNA extraction

Of publicly available accessions of popcorn, 299 unique accessions were selected from the Germplasm Resources Information Network (GRIN), and 21 accessions were selected from the International Maize and Wheat Improvement Center (CIMMYT). Of these, 79 accessions had available sequence information (Romay et al. 2013). Thus, 241 new popcorn accessions were sequenced, with 68 selected for replicate sequencing. Additionally, four dent inbreds with available sequence information were included as outgroup accessions: B73, Mo17, PH207, and Oh43.

Samples were prepared for DNA extraction by bulking tissue from five seedlings of each accession in a 2 mL tube, collected from the youngest leaf of each approximately ten days after germination. DNA was extracted from each sample using a modified CTAB DNA extraction protocol and resuspended in water. DNA was quantified in 96-well plates using PicoGreen (Invitrogen, Carlsbad, CA), and DNA concentrations were normalized to 50ng/µL by diluting with 10 mM Tris, then stored at 4°C.

2.2 Genotyping-by-sequencing library construction

Genotyping-by-sequencing libraries were constructed using the protocol described by Elshire *et al.* (2011). Briefly, restriction digestion and ligation were performed in 96-well plates using *Ape*KI by adding 250 ng of DNA from each sample, 1.5µL of 0.1 µM rare barcoded adapter, and 0.5µL of 10 µM common barcoded adapter per well. Libraries were pooled, size-selected and cleaned with Agencourt AMPure XP beads (Beckman Coulter Inc., Brea, CA), and amplified for 15 cycles using KAPA HiFi HotStart ReadyMix (KAPA BIOSYSTEMS). Following a final purification with Agencourt AMPure XP beads, libraries were quantified using PicoGreen. A 1 ng/µL dilution of the library was made using 10 mM Tris, and 5 µL was run on a 1% agarose gel to verify size selection and primer-dimer removal. Fragment size was estimated using a Bioanalyzer 2100 (Agilent, Santa Clara, CA) High Sensitivity DNA Analysis. Pooled libraries were diluted to 10 nmol using 10 mM Tris and sequenced with 100-bp, single-end reads on an Illumina NovaSeq 6000 (Illumina, San Diego, CA). The sequencing data has been submitted to NCBI SRA (PRJNA911304).

2.3 Marker calling

The reads from the 79 popcorn accessions previously sequenced by Romay et al. (2013) were downloaded from NCBI SRA (SRP021921) and processed with the new raw sequencing reads using the GBS v2 discovery pipeline from TASSEL 5 (release 5.2.59) (Glaubitz et al. 2014). The pipeline tools and select parameters were run in the following order: First, GBSSeqToTagDBPlugin, with ApeKI a kmer length of 64, and a minimum read quality score of 20, followed by TagExportToFastqPlugin. Next, two alignment indices were built with Bowtie2 (Langmead & Salzberg 2012) using the B73 v5 and HP301 v1 reference genomes (Hufford et al. 2021). Both genome indices and output files were processed using the same parameters afterward. The Bowtie2 alignment command was run using the --very-sensitive preset, resulting in an overall alignment of 90.6% and 91.2% for the B73 v5 and HP301 v1 indices, respectively. The SAM files were then sorted using samtools (Li et al. 2009), and

mapping quality (MAPQ) scores were analyzed using Qualimap 2 (Okonechnikov et al. 2016). The SAM files were read using the SAMToGBSdbPlugin with a minimum MAPQ score of 20, after which the DiscoverySNPCallPluginV2 was run. Finally, the ProductionSNPCallerPluginV2 was run using the same parameters as the GBSSeqToTagDBPlugin.

2.4 Marker coverage and characterization

The genomic distribution of single nucleotide polymorphisms (SNPs) was based on the HP301 v1 reference genome, as the popcorn reads aligned slightly better than when using the B73 v5 genome (see above). Markers were reported from TASSEL 5 for the Core Analysis marker set and each chromosome using a minimum minor allele frequency of 0.01. To assess inbreds and heterogeneous accessions separately, the Core Analysis marker set was filtered to exclude polymorphic sites with more than 0.9 missing data. A filtered marker set was generated for imputation and GWAS after removing sites with minor allele frequencies less than 0.05 and more than 0.75 missing data. Unless otherwise stated, the Core Analysis marker set was used for all analyses. These marker sets are described in detail below (Section 3.2) and are available for public download at https://datacommons.cyverse.org/browse/iplant/home/msullivan/Rev-2023-Popcorn-Structure-Diversity. Markers were identified as intragenic by searching their positions relative to the start and stop positions of the HP301 v1 gene models. The density and distribution of polymorphic SNPs were determined in *R* (R Core Team, 2021) using the *R* package *rMVP* (Yin et al. 2021).

2.5 Population structure

A kinship matrix of centered identity by state (IBS) values was calculated for all possible pairwise taxa comparisons, within heterogeneous and inbred accessions separately, from the Core Analysis

marker set in TASSEL. Plots of each pairwise comparison were generated using the *R* package *ggplot2* (Wickham 2016). To formally test for population structure using molecular data, unsupervised model-based clustering analysis was performed on inbred accessions, heterogeneous accessions, and the combined dataset using ADMIXTURE (Alexander et al. 2009). Heterogeneous and inbred accessions were tested using one through five *K* populations and tenfold cross-validation, with errors minimized at *K* equals two for heterogeneous accessions and *K* equals four for inbred accessions. The combined dataset tested one through nine *K* populations and tenfold cross-validation, with errors minimized at *K* equals two. Individuals were assigned to a group only if assignment was greater than 0.6. An additional ADMIXTURE analysis was performed in validation after combining the 320 popcorn accessions with 2,284 additional accessions sequenced from the Ames Diversity Panel (Romay et al. 2013). Due to computational restraints, accessions designated stiff stalk or non-stiff stalk were excluded. Reads from the combined set of accessions were aligned to the HP301 v1 reference genome using the TASSEL 5 GBS v2 discovery pipeline without changing any of the aforementioned parameters. In ADMIXTURE, one through twenty *K* populations were tested using fivefold cross-validation, with errors minimized at *K* equals fifteen.

Principal coordinates were generated in TASSEL from each of the IBS kinship matrices produced from the three datasets. Two coordinates from each matrix were visualized using *ggplot2*. Phylogenetic analysis was performed using the *R* package *ape* (Paradis et al. 2004) using the neighbor-joining method from a modified Euclidean distance matrix and visualized with the *R* package *ggtree* (Yu et al. 2017). The neighbor-joining algorithm was selected because of the considerable differences in selection and breeding history of the populations included in this study, making alternate methods such as UPGMA untenable. To further investigate the relationship between teosinte and the Pointed and Latin American Popcorns, an additional neighbor-joining tree was generated. A distance matrix

was produced from the seventeen teosinte accessions and ten randomly selected individuals from each of the fifteen ADMIXTURE groups identified from the combined popcorn – Ames Diversity Panel analysis. This distance matrix was used by *ape* to create the second tree and also visualized with *ggtree*.

2.6 Population genetics and diversity

To evaluate differences in selection and genetic drift between heterogeneous and inbred accessions, the proportion of monomorphic SNPs within each group was calculated. Each group was filtered from the Core Analysis marker set, after which markers were filtered to contain a minimum of two observations. Markers without minor alleles were then filtered out, allowing the number of monomorphic markers to be calculated by subtracting the remaining polymorphic markers. After filtering monomorphic sites, marker minor allele frequencies were calculated from the Core Analysis marker set using the TASSEL 5 GenotypeSummaryPlugin. The distribution of minor allele frequencies was calculated for each popcorn group.

A marker subset was produced for linkage disequilibrium (LD) analysis by removing taxa with greater than 0.8 missing data, SNPs with more than 50% missing data and heterozygosity greater than 0, resulting in 29,161 retained markers. Within each group, LD was calculated between the subset of markers on each chromosome using the TASSEL 5 LinkageDisequilibriumPlugin. The level of LD decay was measured using the squared Pearson correlation coefficient (r^2) between each pair of markers. Pairwise comparisons were binned according to distance using logarithmically increasing bins.

2.7 Herbicide treatment

A subset of popcorn accessions was evaluated for response to nicosulfuron at the University of Illinois Crop Sciences Research and Education Center, Urbana, Illinois, during the summer of 2022. A

total of 294 genotypes were included in the trial, with 111 North American Yellow Pearl Popcorns and 178 Pointed and Latin American Popcorns. Additionally, five controls were included in the panel – a tolerant sweet corn hybrid (GSS1477), a nicosulfuron-sensitive sweet corn hybrid (Merit), a partially sensitive yellow popcorn hybrid, a tolerant yellow popcorn hybrid, and a tolerant white popcorn hybrid. Genotypes were arranged in a randomized complete block design with three replications. An experimental unit comprised a single 3-meter row containing 10-20 plants.

When the majority of plants reached the V3/V4 growth stage, nicosulfuron was applied to one-half of each row at a rate of 35 g ai ha^{-1} with a 0.25% (v/v) nonionic surfactant and 2.5% (v/v) urea ammonium nitrate. This treatment was delivered using a pressurized CO_2 backpack sprayer equipped with AI110025 nozzles (TeeJet Technologies, P.O. Box 7900, Wheaton, IL 60187) spaced 51 cm apart on a 1.5-meter boom calibrated to apply 187 L ha^{-1} at 276 kPa. Popcorn plants were evaluated ten days after nicosulfuron application. Injury was assessed visually on a scale of 0-4, with 0 representing no injury and 4 representing plant death.

2.8 Imputation and genome-wide association studies (GWAS)

To evaluate the imputation accuracy of different methods, 1% of the GWAS Filtered set was masked and imputed. Imputation was performed with FILLIN (Swarts et al. 2014) and Beagle v5.1 (Browning et al. 2018). FILLIN was run using a haplotype block size of 4,096. Beagle was run using an effective population size of 100,000 and used a genetic map generated from the B73 x HP301 NAM population. The genetic map for HP301 was constructed in *R/qtl* (Broman et al. 2003) using GBS data from the HP301 RILs (Elshire et al. 2011), aligned to the HP301 v1 reference genome using the TASSEL 5 GBS v2 discovery pipeline. To compare the two imputation methods, the unmasked,

masked, and imputed marker sets were used as input for TASSEL's ImputationAccuracyPlugin, and error rates were generated.

As the popcorn groups differ in kernel color and response to nicosulfuron, GWAS was performed on both these phenotypes. Refined phenotypes for the GWAS were generated by using the median nicosulfuron injury and average endosperm color reported from GRIN. Both GWAS analyses were performed using a weighted MLM in TASSEL with the FILLIN-imputed marker set, a kinship matrix of normalized IBS values, and a weight matrix of principal coordinates with high eigenvalues. A Bonferroni-corrected p-value of 0.01 resulted in a -log(p) significance threshold of 7.3. Results of both GWAS analyses were visualized using the *R* package *qqman* (Turner 2018).

3 RESULTS

3.1 Germplasm

To improve the genomic resources and investigate the genetic diversity of popcorn, GBS was performed on 320 popcorn accessions. In GRIN, a total of 640 available accessions were listed as popcorn. Accessions that were also labeled as other kernel types were removed, reducing the count to 379. Then all accessions that appeared to be duplicates based on passport information were removed, bringing the number of unique popcorn accessions down to 304. Of the accessions with multiple kernel types listed, 37 were added back, as they were known popcorn lines. Of these 341 accessions, 42 were removed due to poor sequencing. Finally, an additional 21 popcorn landraces from CIMMYT were included, bringing the total number of accessions to 320 (Supplementary Table S1). Of these, 148 are described as partially or highly inbred and 172 as populations or landraces. To understand the relatedness and placement of North American popcorn germplasm relative to other groups of popcorn, including non-inbred accessions was necessary, as most publicly available

popcorn accessions outside of the United States are not inbred. Additionally, most inbred popcorn accessions in GRIN are derived from North American Yellow Pearl breeding programs, so heterogeneous accessions, including the populations and landraces, were included to increase the total number of accessions belonging to previously reported groups of popcorn, such as the North American Early or Pointed Popcorns. In addition to the 320 popcorn accessions, four previously sequenced dent inbreds were included to serve as outgroup lines: B73, Mo17, PH207, and Oh43.

3.2 Marker density and distribution

From the GBS data, a Core Analysis marker set of 308,811 markers with an average minor allele frequency of 0.17228 was generated after passing SNP filtering and quality control. Of the Core Analysis marker set, 250,966 (81.3%) were identified as intragenic using the HP301 v1 gene models (Hufford et al. 2021), consistent with the observation that SNP density increased from the centromeric to the telomeric regions (Supplementary Fig. S1). Across the 10 maize chromosomes, the Core Analysis marker set had an average marker density of 146 SNPs per Mb, with the highest and lowest marker densities observed on chromosomes 5 (159.8 SNPs per Mb) and 6 (136.7 SNPs per Mb), respectively (Table 1). A Filtered GWAS set of 200,101 markers with an average minor allele frequency of 0.216 was generated by removing all SNPs with a minor allele frequency less than 0.05 and greater than 0.75 missing data in preparation for imputation. The Filtered GWAS marker set had an average marker density of 94.6 SNPs per MB, with the highest and lowest densities on chromosomes 1 (108 SNPs per Mb) and 4 (74.7 SNPs per Mb), respectively.

Table 1 SNP characteristics summary for the Core Analysis and Filtered GWAS SNP marker sets

		Core Analysis Marker Set			Filtered GWAS Marker Set		
Chromosome	Chromosome Length (b)	Number of SNPs	SNPs per Mb	MAF	Number of SNPs	SNPs per Mb	MAF
1	305,741,490	48164	157.5	0.17085	32903	107.6	0.21527
2	245,520,863	37750	153.8	0.17896	24514	99.8	0.22231
3	240,329,036	35548	147.9	0.17618	24554	102.2	0.22337
4	251,015,817	29831	118.8	0.16678	18748	74.7	0.20173
5	221,904,911	35466	159.8	0.17589	23121	104.2	0.22609
6	177,672,499	24287	136.7	0.16776	14358	80.8	0.21071
7	180,265,189	26651	147.8	0.16964	15815	87.7	0.21834
8	177,329,642	25532	144.0	0.16633	16579	93.5	0.2199
9	164,337,916	23848	145.1	0.17361	16053	97.7	0.21045
10	150,435,773	21734	144.5	0.173	13456	89.4	0.21866
Total	2,114,553,136	308811	146.0	0.17228	200101	94.6	0.21576

3.3 Population structure

Although there is value to including heterogeneous accessions along with inbred accessions, direct comparison of some measures of diversity and population may be inappropriate, as heterozygosity and heterogeneity will affect those measures, resulting in exaggerated differences between heterogeneous and inbred accessions. Thus, for several analyses, heterogeneous and inbred accessions were analyzed separately, with a combined dataset of both types of accessions when appropriate. To identify relationships among accessions included in the panel, the distribution of IBS

pairwise comparisons for the 320 accessions was calculated using the Core Analysis marker set (Fig. 1). Heterogeneous accessions showed a normal distribution centered on 0.71 (Fig. 1A). However, within inbreds, a bimodal distribution was identified, with a two peaks centered on 0.69 and an additional smaller peak centered on 0.82 (Fig. 1B). These results indicate that while most heterogeneous accessions in this panel have a weak relationship with each other, the inbred accessions appear to contain at least two distinct groups.

The North American popcorn populations are comprised of accessions with different breeding histories, geographic distributions, inbreeding status, and selection pressures. An ADMIXTURE analysis was performed on heterogeneous and inbred accessions separately, as well as a combined set of all taxa, to formally test for population structure using one to six *K* populations.

When analyzing heterogeneous accessions, cross-validation error was minimized at *K*=2 (Supplementary Fig. S2A), and the 172 accessions were clustered into two populations, with 28 individuals assigned to the first and 131 assigned to the second, leaving 13 admixed accessions unassigned (Fig. 2A, Supplementary Table S2). Most taxa assigned to the first group have round yellow kernels and can be described as North American Yellow Pearl Popcorns, including several populations from breeding programs in the United States with known pedigrees as well as the founders of the previously described subgroups, Amber Pearl, Supergold, and South American. Interestingly, two accessions from Mexico and Brazil with round yellow kernels were assigned to this group – SONO 147 and BRAZ 2832. Additionally, two notable taxa from Chile were assigned to this group, CHZM 06 004 and CHZM 06 012, both of which are members of the Curagua race. The remaining members of Curagua, except one which was unassigned, were assigned to the other group of popcorn, which will be referred to as the Pointed and Latin American Popcorns. This group contains many diverse accessions, including all North American Pointed Rice and Early Popcorns

(n=21), all Latin American pointed and pearl popcorns accessions outside the four accessions described above (n=75) and nearly all accessions developed outside of the Americas (n=35).

When only considering the 148 inbred accessions, cross-validation error was minimized at *K*=4 (Supplementary Fig. S2B), with 22 individuals assigned to the first, 24 assigned to the second, 26 assigned to the third, and 36 assigned to the fourth, leaving 40 admixed accessions unassigned (Fig. 2B, Table S3). All taxa assigned to the first and second groups are members of the previously described South American and Supergold subgroups, respectively. The third group is mostly comprised of other North American Yellow Pearl Popcorns, many of which have introgressions from dent and flint material. The fourth group is analogous to the Pointed and Latin American Popcorns, and contains all pointed inbreds, inbreds derived from landraces from Latin America, and several inbreds derived from the North American Early Popcorn Tom Thumb.

When all taxa were assessed together, cross-validation error was again minimized at *K=2* (Supplementary Fig. S2C), with a group corresponding to the North American Yellow Pearls (n=135), and another to the Pointed and Latin American Popcorns (n=170), leaving 15 accessions unassigned (Fig. 2C, Supplementary Table S4).

A second analysis was performed to further examine the population structure in popcorn and to ensure that the identification of additional popcorn groups was not limited by an insufficient number of related accessions. A combined dataset was generated after merging the heterogeneous and inbred popcorn accessions with 2,284 accessions from the Ames Diversity Panel that were not categorized as dent (stiff stalk or non-stiff stalk). After minimizing the ADMIXTURE cross-validation error at *K*=15, the combined 2,604 accessions were assigned to 15 groups, with two popcorn groups identified, corresponding to the North American Yellow Pearl Popcorns and the Pointed and Latin

American Popcorns. Although dent lines were excluded, 885 unclassified accessions were assigned to known public subgroups, including B73, B14, B52, Hy, Mo17, W22, Oh43, and WF9, demonstrating that dent material was still well-represented in the analysis. Notably, the analysis assigned the 17 teosinte inbred lines included in the Ames Diversity Panel to the Pointed and Latin American Popcorn group. Thus, the teosinte inbred lines are most genetically similar to the Pointed and Latin American Popcorns, followed by North American Yellow Pearl Popcorns, then sweet corn, and finally dent and flint tropical and temperate material (Supplementary Fig. S3). Other notable taxa assigned to the Pointed and Latin American Popcorn group include accessions of Palomero Toluqueño, Reventador, Chapalote, Pororo, Avati Pichinga, and Pisankalla.

Principal coordinate analysis (PCoA) was also performed to investigate population structure and identify individual taxa groupings. Heterogeneous and inbred accessions were each assessed independently, as well as together in the combined dataset (Fig. 3). Across all three analyses, the first principal coordinate corresponded to the division between North American Yellow Pearl Popcorns and the Pointed and Latin American Popcorns, supporting the general *K*=2 optimum number of clusters. Of the total variation explained within the ten principal coordinates tested, the first three explained 39.1%, 59.5%, and 58.9% for heterogeneous, inbred, and combined accessions, respectively. Within the heterogeneous and the combined analyses, the second coordinate separates the Pointed and Latin American Popcorns into three rough divisions (Fig. 3A, 3C). At values greater than 0.05, most lowland tropical popcorns, such as Pororo, Avati Pichinga, Argentine Pop, and Pichinga cluster together. Between 0.05 and -0.05, the next cluster is composed of mostly pointed or highland landraces from Latin America, although notable additions include Curagua, Palomero Toluqueno, and Strawberry. The final cluster formed below -0.05 is most apparent in the combined dataset, as it contains all inbred temperate pointed rices, as well as populations such as

Japanese Hulless, Pinky Popcorn, and Spanish Pop. The inbred analysis exhibited separation for each of the ADMIXTURE defined populations, with an additional cluster located at the bottom of each graph. This cluster is comprised of individuals developed by Iowa State's Popcorn Breeding Program and derived pigmented accessions, including IA61, I-28, B-28, and R-28. In the combined analysis of heterogeneous and inbred accessions, not only are the three divisions of North American Yellow Pearl Popcorn distinct, but this subgroup of Iowa popcorns also clusters separately near the top of the North American Yellow Pearl Popcorns, despite not being identified as significant in the ADMIXTURE analysis.

The neighbor-joining tree based on the distance matrix supported the population structure findings of the ADMIXTURE analyses and PcoA (Fig. 4). The major division of clades within the tree corresponded to the separation between North American Yellow Pearl Popcorns and Pointed and Latin American Popcorns. The two Chilean Curagua accessions fall within and basal to the North American Yellow Pearl clade, while the remainder are located within the Pointed and Latin American Popcorns. The two Latin American taxa assigned to the North American Yellow Pearl Popcorn population, BRAZ 2832 and SONO 147, occur between the Supergold and South American clades, indicating that they likely represent a recent movement of North American Yellow Pearl Popcorn germplasm (Fig. 4).

3.4 Population genetics and allele diversity

To evaluate differences in allele frequencies and genetic diversity within popcorn germplasm, each population was evaluated for the proportion of monomorphic SNPs, minor allele frequencies (MAF), and LD decay. After filtering markers in each group to account for structural variation, monomorphism was evaluated. The proportion of monomorphic SNPs was considerably different

between groups, with 2.4 times more monomorphic SNPs in heterogeneous North American Yellow Pearl Popcorns (0.673) than in the Pointed and Latin American Popcorns (0.275). This difference between the two populations was also observed in the inbred accessions, with the highest frequency of monomorphism observed in South American (0.72), followed by Supergold (0.59), other North American Yellow Pearl Popcorns (0.51), and finally the Pointed and Latin American Popcorns (0.16).

Additionally, there were differences in MAF across populations (Fig. 5). Within heterogeneous accessions, the relative frequency of SNPs with a MAF between 0.01 and 0.1 in North American Yellow Pearl Popcorns and Pointed and Latin American Popcorns were 3.5% and 17.5%, respectively (Fig. 5A). The relative frequencies of SNPs at other MAFs were similar between the two heterogeneous groups, although the North American Yellow Pearl Popcorns had slightly greater relative frequencies at MAFs between 0.3 and 0.5. Generally, all inbreds had similar distributions of MAF, with most minor alleles occurring at frequencies of 0.2 or lower, with the Supergolds exhibiting the most extreme example (Fig. 5B). Overall, differences in inbreeding status had a greater effect on MAF than differences between the populations.

Linkage disequilibrium decay differed between the populations (Fig. 6). While median linkage disequilibrium decay was comparable within heterogeneous accessions (Fig. 6A), the inbreds exhibited considerable differences (Fig. 6B). The South American and Supergold subpopulations experienced the slowest LD decay, followed by the other North American Yellow Pearls. However, LD in inbred Pointed and Latin American Popcorns decayed nearly as fast as the heterogeneous germplasm. When compared with previous evaluations of LD decay in temperate and tropical germplasm (Romay et al. 2013, White et al. 2020), inbred North American Yellow Pearl Popcorns exhibit slower LD decay than even the B73 heterotic subgroup (r^2 of ~0.25 within 10 – 99 Kb), while

LD decay in the inbred Pointed and Latin American Popcorns was comparable to that of CIMMYT tropical inbreds (r^2 of 0.1 within 100 – 999 bp).

3.5 Genome-wide association studies

A significant limitation of GWAS is the *p*-value correction required to account for multiple tests performed (Tam et al. 2019). Reducing the number of markers tested is one way to ease the stringency of the commonly used Bonferroni correction method. By removing all markers from the Core Analysis marker set with minor allele frequencies less than 0.05 and missing data greater than 0.75, the resulting Filtered GWAS set reduced the number of markers to approximately half of the Core Analysis marker set. Furthermore, since the imputation strategies tested here rely on generating haplotypes from existing marker data, any reduction in missing sites improves haplotype generation and subsequent imputation. Two commonly used imputation strategies in maize are Beagle and FILLIN. Both were used to impute the Filtered GWAS set, and error metrics were calculated for each. FILLIN and Beagle had overall error rates of 0.02321 and 0.08173, respectively. Additionally, while imputation with Beagle left no missing data, imputation with FILLIN decreased missing data in inbred accessions from 0.6 to 0.3 and in heterogeneous accessions from 0.91 to 0.84. Because of the difference in error rates, only the FILLIN-imputed marker set was used to perform GWAS.

The North American Yellow Pearl Popcorns and Pointed and Latin American Popcorns represent two distinct groups with significant differences in population structure, genetic diversity, and breeding history. Additionally, phenotypic differences are known between these two groups, such as kernel color. North American Yellow Pearl Popcorns are typically yellow, with only four white accessions in the panel (3.3%), while the Pointed and Latin American Popcorns have a much higher frequency of

white kernels or endosperm, with 111 accessions in the panel (54.7%). Since kernel color is a simple trait that differs among the germplasm assessed here, it was selected as a proof of concept to assess the utility of using this panel for GWAS (Fig. 7, Supplementary Fig. S4). Several significant SNPs were identified, some of which have been associated with kernel color previously (Supplementary Table S5). The SNP with the strongest association with kernel color is located in *Y1*, which is known to affect the concentration of carotenoid pigments in the endosperm. The SNP with the next strongest association with kernel color is located 200 bases away from Zm00027ab258670, a gene model predicted to have carotene desaturase activity. Three other peak SNPs were identified, none of which have been previously associated with kernel color. These results indicate that despite the known differences in population structure and phenotype frequencies, the panel can identify associations between loci and traits.

A trait anecdotally associated with kernel color in popcorn is herbicide injury. Herbicide labels often exclude white popcorn from approved use, even when yellow popcorn is included (Loux et al. 2017). While the white color itself should have no bearing on injury, herbicide sensitivity differences between the two groups may have been reduced to kernel color for simplicity. In the herbicide trial, 4 of the 111 North American Yellow Pearl Popcorns (3.6%) and 32 of 178 Pointed and Latin American Popcorns (18%) showed typical nicosulfuron injury (Supplementary Table S6). Of the 32 injured Pointed and Latin American Popcorns, several are important historical accessions grown in the American Midwest, including Japanese Hulless, Australian Hulless, H5505, IA DS 38-W, and IA DS 43-W. Notably, these pointed accessions are all white and represent much of the variation of the pointed rice popcorns traditionally grown in North America.

The Filtered GWAS set was used to map nicosulfuron injury. A total of nine SNPs representing seven peaks were found to be significant (Fig. 8, Supplementary Fig. S5, Supplementary Table S7). The SNP

with the strongest association with injury is located in *cl11433_2b*, a transcription factor associated with the thylakoid. The next strongest association was with a subunit of NADH dehydrogenase. Other SNPs with statistically significant associations include a predicted protein kinase, the vacuole-associated phospholipase *pco139881*, a locus 1.3 kb upstream from a predicted aminoacyl-tRNA synthetase, the cytochrome p450 *CYP72A27*, and a predicted monooxygenase glycosyltransferase. Although these genes have not previously been implicated in nicosulfuron injury, many of the classes to which they belong are associated with herbicide tolerance.

4 DISCUSSION

4.1 Population Structure

The results of IBS, ADMIXTURE, PCoA, and the neighbor-joining tree support a major division within the popcorn included in this panel: the North American Yellow Pearl Popcorns and the Pointed and Latin American Popcorns. Additionally, significantly different subgroups within the North American Yellow Pearl Popcorns were identified, which largely separated two heterotic pools from other North American Yellow Pearl Popcorns: Supergold and South American popcorn. Both the PcoA and neighbor-joining tree indicate a degree of sub-grouping within the Pointed and Latin American Popcorns, but these subgroups were not significantly different based on the genomic data. Three general subgroups were identified within the Pointed and Latin American Popcorns: North American Pointed Rice Popcorns, Latin Pointed Rice and highland popcorn landraces, and lowland tropical popcorn landraces. It is important to note that not all individuals typically assigned to one of these subgroups clustered with their subgroup, though the majority do. Although subgroups were identified in the Pointed and Latin American Popcorns, they are not comparable in magnitude to those of the North American Yellow Pearl Popcorns included in this panel. Therefore, a single major

division is proposed here that, along with subgroups, explains most of the variation in population structure in this popcorn diversity panel.

Foundational work on the phylogenetic relationships of North American popcorn identified several groups of popcorn. After evaluating 56 popcorn accessions for morphological traits, isozymes, and SSRs, six major divisions were previously identified (Santacruz-Varela et al. 2004). Two subgroups of pointed rice popcorn of North American and Latin American origin were proposed, followed by the related North American Early Popcorns. A single group of North American Yellow Pearl Popcorns was identified, with minimal divisions between the heterotic groups. Finally, two additional groups of Latin America were identified, one corresponding to lowland races of Argentina, Brazil, and Paraguay, and the other composed of various accessions from Chile, Uruguay, and Mexico.

Besides the number of accessions, a major difference between previous studies of North American Popcorn germplasm and the results presented here is that morphological traits were not included when determining phylogenetic relationships and population structure. All traits are composed of genotypic and environmental effects, and many of the vegetative and phenological traits included in previous analyses of popcorn are known to be affected by photoperiod sensitivity, including plant height, ear height, leaf number, and days to flowering (Fei et al. 2022). As few as four genomic regions have been identified that control photoperiod responses in these traits (Coles et al. 2010). Therefore, including these traits in measures of phylogenetic divergence likely does not reflect genomic differences and could be heavily influenced by photoperiod sensitivity. This is supported by the conclusion that the six previously identified popcorn groups were primarily assigned based on results of the morphological characterization (Santacruz-Varela et al. 2004). The inclusion of morphological traits that do not reflect genome-wide differences may explain the discrepancy between the six groups of popcorn identified previously and the groups presented here.

Investigation into the genetic diversity and population structure of CIMMYT accessions and Chinese germplasm has identified distinct groups of popcorn in their panels. A panel composed of teosinte, popcorn landraces, non-popcorn landraces, North American Yellow Pearl Popcorns, and elite CIMMYT inbreds was used to identify loci associated with temperate and tropical adaptation (Li et al. 2019). This panel identified two groups of popcorn, one corresponding to the North American Yellow Pearl Popcorns and the other associated with landraces from Latin America. The genetic diversity and population structure within a popcorn collection of Chinese accessions, CIMMYT landraces, and inbreds from the United States were recently characterized (Yu et al. 2021). In this study, popcorn accessions were assigned to one of three groups – Latin American landraces, North American Yellow Pearl Popcorns, and recently developed inbreds from breeding programs, likely of North American Yellow Pearl origin. Based on these studies and the findings presented here, North American Yellow Pearl Popcorns have consistently been shown to constitute a genetically distinct pool of germplasm compared to Latin American landraces.

In contrast with the uniform separation of North American Yellow Pearl Popcorns from other types of maize, the assignment of the Pointed and Latin American Popcorns is not as consistent. Single-group assignment of popcorn landraces of Latin America has been previously reported when compared with Chinese germplasm as discussed above (Yu et al. 2021), Brazilian popcorn inbreds (Vittorazzi et al. 2018), and highland and lowland Argentine landraces (Bracco et al. 2012). When compared with other landraces with geographic proximity, most popcorn landraces across the Americas cluster as a single group (van Heerwaarden et al. 2011). However, a more recent investigation of Latin American and Caribbean landraces identified sixteen groups, three of which included popcorn (Bedoya et al. 2017). The first of the three groups contained Palomero Toluqueño, Chapalote, and Reventador, all of which are considered primitive popcorns of Northern Mexico. The

second group was related to the first and contained Canguil, Confite Puneño, and Pisankalla, all of which are pointed. The third group was unrelated to the first two and contained lowland Avati popcorn. More recently, an analysis of 575 popcorn landraces from Latin America identified nine genetic clusters using discriminant analysis of principal components (DAPC) without considering a priori groups (De Almeida Silva et al. 2020). However, it is important to consider differences in methods when comparing these findings.

Most popcorn groups have been identified using model-based clustering approaches such as STRUCTURE (Pritchard et al. 2000) or ADMIXTURE. Furthermore, model-based clustering methods assign individuals to groups relative to the entire panel – populations identified by an initial clustering analysis may be further divided into subgroups by subsequent clustering analyses, as with the American and Caribbean landraces discussed above. Additionally, methods based on PCA, like DAPC, tend to provide higher estimates of the total number of groups than model-based clustering approaches (Intarapanich et al. 2009, Linck & Battey 2019) and are especially sensitive to the influence of a priori vs. de novo groupings (Miller et al. 2020). Because of these considerations, assignment by the model-based clustering method ADMIXTURE on the Core Analysis marker set was used to assess population structure for the entire panel. Additionally, due to differences in inbreeding status and population development, heterogeneous accessions were separated from inbreds, and each group of accessions was assigned to clusters with ADMIXTURE independently. The findings reported here largely agree with similar approaches to estimate population structure in popcorn, and differences with other reports largely stem from assignment methods.

A limitation in this study is the bulking of tissue from several individuals that constituted a sequenced sample. Bulking tissue in maize has been shown to be a time- and cost-effective strategy for large-scale genotyping in open-pollinated varieties of maize (Reif et al. 2005, Li et al. 2019, De

Almeida Silva et al. 2020, Yu et al. 2021). More recently, large-scale genotyping-by-sequencing of inbreds has involved bulking leaf tissue prior to DNA extraction (Wu et al. 2016, Li et al. 2021). Although commonly employed to represent the diversity within a population, bulking can result in unequal DNA contributions of individuals, introduce false variation in the event of contamination, or overrepresent residual heterozygosity. Additionally, especially in the case of bulked landraces, haplotype-based methods of imputation, including Beagle and FILLIN, struggle to impute, as two or more haplotypes may be present in any given genomic region. Despite this limitation, the results presented here are consistent across analyses, identifying a major division within the panel of two groups of popcorn that exhibit genetic and phenotypic differences.

4.2 Genetic diversity

Heterogeneous and inbred Pointed and Latin American Popcorns contain a variety of accessions, including pointed inbreds from the American Midwest. It is unlikely that the genetic similarities between pointed popcorns from North America and Latin America indicate a recent introduction or transfer of germplasm between continents, as accessions constituting this group were geographically widely distributed. Instead, these similarities may indicate a pre-Columbian interchange of popcorn between North and South America, as has been previously suggested (Bedoya et al. 2017). Additionally, the genetic similarity may reflect reduced admixture with other landraces due to the high frequency of the dent-sterile *Ga1-S* allele in North and South American popcorn (Goodman et al. 2021). This aligns with the finding that popcorn landraces tend to group separately from other maize types, as discussed above.

North American Yellow Pearl Popcorns share genetic similarities with several Curagua accessions from Chile. Genetic similarity between United States varieties and South American material has been

noted before (Vigouroux et al. 2008, van Heerwaarden et al. 2011). However, this has generally been attributed to the post-Columbian movement of germplasm from the United States to South America. Many Chilean, Argentine, and Brazilian races are of United States origin, including Dulce Golden Bantam, Hickory King, and Dente Branco. The presence of Curagua accessions in the North American Yellow Pearls is evidence that intercontinental movement from South to North also occurred.

The North American Yellow Pearl Popcorns have a much higher LD and proportion of monomorphic markers than Pointed and Latin American Popcorns and other groups of maize (Romay 2013). North American Yellow Pearl Popcorns also exhibit higher IBS values when compared with other North American Yellow Pearl Popcorns, indicating a high rate of inbreeding and lower genetic variation. Considering the Curagua accessions clustered among the North American Yellow Pearls (Fig. 4), this reduced genome variation is consistent with a small founder population and fewer meiotic events relative to the considerably older Pointed and Latin American Popcorns. The genetic variation within accessions of Curagua assigned to the North American Yellow Pearls represents only a portion of the total variation present in all members of the Curagua race. If members of Curagua were transported to New England by sailors, as previously suggested (Smith 1999), they would have likely contained a limited gene pool, and subsequent inbreeding beginning in the 1940s would have exacerbated the limited genetic diversity. However, it appears that after development in the United States, some of this germplasm returned to Latin America, as evidenced by the SONO 147 and BRAZ 2832 accessions (Fig. 4). Although little additional information is known about BRAZ 2832, the passport data for SONO 147 (National Genetic Resources Program, 2022) contains the alternate name “Reventador Gringo”, suggesting that growers associated this accession with non-native material.

4.3 Kernel color, popcorn groups, and nicosulfuron injury

The product of the *Y1* gene catalyzes the first step in the maize carotenoid biosynthesis pathway. A SNP in this gene has been identified as a causal mutation that results in a Thr to Asn substitution (Fu et al. 2013). The peak SNP identified in the GWAS is tightly linked with the causal SNP, separated by 346 bases. Another SNP occurs close to a gene model predicted to have carotene desaturase activity and is located 3 Mb away from a locus previously associated with popcorn flake color in CIMMYT germplasm (Li et al. 2021). Given the structural variation between popcorn germplasm and most material used to evaluate kernel color, it is plausible that additional loci may be identified outside the core genome (Lu et al. 2015, Hufford et al. 2021). The three remaining SNPs reported here have not been previously associated with kernel color, although two occur in known genes, *wrky11* and *nc3*, and the other in an uncharacterized gene. Further investigation is necessary to determine whether these additional associations represent biologically meaningful relationships or spurious correlations from population structure.

Recent studies have been unable to identify differences in herbicide sensitivity among several (i.e., up to eight entries) commercial white and yellow popcorn hybrids (Barnes et al. 2020a; Barnes et al. 2020b). However, given the unknown origin of the hybrid parents used in these studies and the prevalence of North American Yellow Pearl Popcorns in commercial germplasm, these hybrids were likely selected using kernel color alone and do not represent the two groups of North American popcorn. The results presented here show significant differences in the two popcorn groups for sensitivity to nicosulfuron. Kernel color has no bearing on herbicide injury and demonstrates the importance of population structure in genetic analyses. Kernel color represents a major phenotypic

difference between the generally nicosulfuron-tolerant North American Yellow Pearl Popcorns and the more susceptible Pointed and Latin American Popcorns.

Nicosulfuron, along with other sulfonylurea herbicides, inhibits acetolactate synthase (ALS), the first enzyme in the biosynthesis of branched-chain amino acids (BCAAs). Although encoded by the nuclear genome, ALS is localized in the chloroplast (Mazur et al. 1987). Along with other enzymes involved in BCAA biosynthesis, ALS interacts with many substrates and products of cellular respiration, including pyruvate, NAD, and FAD (Duggleby et al. 2008). While target-site resistance of sulfonylurea herbicides has been demonstrated in maize (Anderson & Georgeson 1989; Li et al. 2020), non-target site resistance is more common and complex (Délye 2013). Non-target site resistance involves initial detoxification, often in the form of cytochrome p450s and other oxidases, followed by conjugation of thiols or glucosyl groups, and subsequent transport and degradation in the vacuole or extracellular spaces (Yuan et al. 2007).

In sweet and dent corn, detoxification of several post-emergent herbicides, including nicosulfuron, is driven largely by *CYP81A9*, a cytochrome p450 produced by the *Nsf1* locus (Nordby et al. 2008, Choe and Williams 2020). Although this does not appear to be the case with nicosulfuron response in popcorn, other candidate genes were identified. Since the pool of available NADH in oxidative phosphorylation is driven largely by pyruvate, variation in NADH dehydrogenase may allow plants to tolerate changes in pyruvate as ALS becomes bound by nicosulfuron. Additionally, variation in glycosylation and changes in the vacuole is consistent with non-target site resistance. Furthermore, altered amino acid synthesis and phosphorylation may be involved in the alternate metabolic flux of some BCAAs (Joshi et al. 2010). Finally, as many of these processes occur within, or are related to, the chloroplast, changes in the transcription factors of chloroplast-associated genes would be expected.

Although *CYP72A27* is a cytochrome p450, its identification here is somewhat unexpected, as it belongs to the CYP clan 72 group of cytochrome p450s. Members of this clan are generally associated with secondary metabolism, while clan 71 enzymes, such as *CYP81A9*, are responsible for herbicide detoxification (Prall et al. 2016; Brazier-Hicks et al. 2022). However, sulfonylurea herbicide metabolism via clan 72 enzymes has been shown in other crops, such as *CYP72A31* in *Oryza indica* (Saika et al. 2014) and *CYP749A16* in *Gossypium hirsutum* (Thyssen et al. 2018). Furthermore, several members of CYP72 were shown to be differentially expressed in nicosulfuron tolerant and susceptible lines of maize when treated with the herbicide (Liu et al. 2015). Alternatively, the role of *CYP72A27* in detoxifying nicosulfuron may be unique to the Pointed and Latin American Popcorns and simply not previously reported. However, it is possible that rather than directly detoxifying nicosulfuron, *CYP72A27* and other genes identified here respond more like safener-induced enzymes, conferring protection via a stress signaling pathway (Riechers et al. 2010). Finally, as with other candidate genes identified using GWAS, *CYP72A27* may have a spurious correlation with nicosulfuron response. Determining the roles and mechanisms of the genes presented here with nicosulfuron tolerance would require further investigation, including the phenotyping and genotyping of additional members of the Pointed and Latin American Popcorns.

5 CONCLUSION

High-quality genetic resources are important tools for maize breeding programs. A collection of 320 popcorn accessions were analyzed here using 308,811 SNPs. Two distinct groups of popcorn were identified that exhibit genetic and phenotypic differences. North American Yellow Pearl Popcorns constitute a pool of germplasm with reduced genetic diversity and greater rates of inbreeding when compared with the Pointed and Latin American Popcorns. Additionally, subgroups of North American Yellow Pearl Popcorns were identified and largely align with known heterotic groups.

Kernel color frequencies were shown to differ between the two groups, along with sensitivity to nicosulfuron. Novel candidate genes responsible for nicosulfuron tolerance were identified in popcorn. The genomic characterization provided here should be incorporated into popcorn breeding programs to direct improvement and accelerate the rate of genetic gain. Additionally, this collection of popcorn germplasm and genetic data can be used by the maize research community to explore the genetic architecture of traits in popcorn and be compared with loci identified in other groups of maize.

ACKNOWLEDGMENTS

The authors thank Nicholas Hausman for managing the field design, planting, and nicosulfuron application, Crookham Company for the popcorn hybrids used in the herbicide trial, and Marlee Labroo, Lucas Roberts, and Lynn Doran for assistance with library preparation.

AUTHOR CONTRIBUTIONS

Madsen Sullivan: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. Martin M. Williams II: Conceptualization, Methodology, Resources, Writing – review & editing. Anthony J. Studer: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, *19*(9), 1655-1664.

Anderson, P. C., & Georgeson, M. (1989). Herbicide-tolerant mutants of corn. *Genome*, *31*(2), 994-999.Li, Y., Zhu, J., Wu, H., Liu, C., Huang, C., Lan, J., ... & Xie, C. (2020). Precise base editing of non-allelic acetolactate synthase genes confers sulfonylurea herbicide resistance in maize. *The Crop Journal*, *8*(3), 449-456.

Barnes, E. R., Lawrence, N. C., Knezevic, S. Z., Irmak, S., Rodriguez, O., & Jhala, A. J. (2020). Dose response of yellow and white popcorn hybrids to glyphosate, a premix of 2, 4-D choline and glyphosate, or dicamba. *Agronomy Journal*, *112*(4), 2956-2967.

Barnes, E. R., Lawrence, N. C., Knezevic, S. Z., Rodriguez, O., Irmak, S., & Jhala, A. J. (2020). Weed control and response of yellow and white popcorn hybrids to herbicides. *Agronomy Journal*, *112*(1), 458-469.

Bautista-Ramírez, E., Santacruz-Varela, A., Córdova-Téllez, L., Muñoz Orozco, A., López-Sánchez, H., & Esquivel-Esquivel, G. (2020). Yield and expansion capacity of the corn grain in the Palomero Toluqueño race. *Revista mexicana de ciencias agrícolas*, *11*(7), 1607-1618.

Bedoya, C. A., Dreisigacker, S., Hearne, S., Franco, J., Mir, C., Prasanna, B. M., ... & Warburton, M. L. (2017). Genetic diversity and population structure of native maize populations in Latin America and the Caribbean. *PloS one*, *12*(4), e0173488.

Bracco, M., Lia, V. V., Hernández, J. C., Poggio, L., & Gottlieb, A. M. (2012). Genetic diversity of maize landraces from lowland and highland agro-ecosystems of Southern South America: implications for the conservation of native resources. *Annals of Applied Biology*, *160*(3), 308-321.

Brazier-Hicks, M., Franco-Ortega, S., Watson, P., Rougemont, B., Cohn, J., Dale, R., ... & Edwards, R. (2022). Characterization of Cytochrome P450s with Key Roles in Determining Herbicide Selectivity in Maize. *ACS omega*.

Broman, K. W., Wu, H., Sen, Ś., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *bioinformatics*, *19*(7), 889-890.

Browning, B. L. (2018). Beagle 5.0. *University of Washington, Division of Medical Genetics*.

Brunson, A. M. (1937). Popcorn breeding.

Buckler, E. S., Holland, J. B., Bradbury, P. J., Acharya, C. B., Brown, P. J., Browne, C., ... & McMullen, M. D. (2009). The genetic architecture of maize flowering time. *Science*, *325*(5941), 714-718.

Choe, E., & Williams, M. M. (2020). Expression and comparison of sweet corn CYP81A9s in relation to nicosulfuron sensitivity. *Pest Management Science*, *76*(9), 3012-3019.

Coles, N. D., McMullen, M. D., Balint-Kurti, P. J., Pratt, R. C., & Holland, J. B. (2010). Genetic control of photoperiod sensitivity in maize revealed by joint multiple population analysis. *Genetics*, *184*(3), 799-812.

Cook, J. P., McMullen, M. D., Holland, J. B., Tian, F., Bradbury, P., Ross-Ibarra, J., ... & Flint-Garcia, S. A. (2012). Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. *Plant physiology*, *158*(2), 824-834.

Da Fonseca, R. R., Smith, B. D., Wales, N., Cappellini, E., Skoglund, P., Fumagalli, M., ... & Gilbert, M. T. P. (2015). The origin and evolution of maize in the Southwestern United States. *Nature plants*, *1*(1), 1-5.

De Almeida Silva, N. C., Vidal, R., Bernardi Ogliari, J., E Costich, D., & Chen, J. (2020). Relationships among American popcorn and their links with landraces conserved in a microcenter of diversity. *Genetic Resources and Crop Evolution*, *67*(7), 1733-1753.

Délye, C. (2013). Unravelling the genetic bases of non-target-site-based resistance (NTSR) to herbicides: a major challenge for weed science in the forthcoming decade. *Pest management science*, *69*(2), 176-187.

Duggleby, R. G., McCourt, J. A., & Guddat, L. W. (2008). Structure and mechanism of inhibition of plant acetohydroxyacid synthase. *Plant Physiology and Biochemistry*, *46*(3), 309-324.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one*, *6*(5), e19379.

Erwin, A. T. (1950). The origin and history of pop corn. *Economic Botany*, *4*(3), 294-299.

Fei, J., Jiang, Q., Guo, M., Lu, J., Wang, P., Liu, S., ... & Guan, S. (2022). Fine Mapping and Functional Research of Key Genes for Photoperiod Sensitivity in Maize. *Frontiers in Plant Science*, *13*.

Fu, Z., Chai, Y., Zhou, Y., Yang, X., Warburton, M. L., Xu, S., ... & Yan, J. (2013). Natural variation in the sequence of PSY1 and frequency of favorable polymorphisms among tropical and temperate maize germplasm. *Theoretical and Applied Genetics*, *126*(4), 923-935.

Gage, J. L., Monier, B., Giri, A., & Buckler, E. S. (2020). Ten years of the maize nested association mapping population: impact, limitations, and future directions. *The Plant Cell*, *32*(7), 2083-2093.

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., & Buckler, E. S. (2014). TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PloS one*, *9*(2), e90346.

Goodman, M. M., Jones, Z. G., Sanchez, G. J., & Kermicle, J. L. (2021). Maize Cross Incompatibility and the Promiscuous Ga1-m Allele. *Plant breeding reviews*, *44*, 31-56.

Hufford, M. B., Seetharam, A. S., Woodhouse, M. R., Chougule, K. M., Ou, S., Liu, J., ... & Dawe, R. K. (2021). De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science*, *373*(6555), 655-662.

Intarapanich, A., Shaw, P. J., Assawamakin, A., Wangkumhang, P., Ngamphiw, C., Chaichoompu, K., ... & Tongsima, S. (2009). Iterative pruning PCA improves resolution of highly structured populations. *BMC bioinformatics*, *10*(1), 1-17.

Joshi, V., Joung, J. G., Fei, Z., & Jander, G. (2010). Interdependence of threonine, methionine and isoleucine metabolism in plants: accumulation and transcriptional regulation under abiotic stress. *Amino acids*, *39*(4), 933-947.

Kantety, R. V., Zeng, X., Bennetzen, J. L., & Zehr, B. E. (1995). Assessment of genetic diversity in dent and popcorn (Zea mays L.) inbred lines using inter-simple sequence repeat (ISSR) amplification. *Molecular breeding*, *1*(4), 365-373.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, *9*(4), 357-359.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079.

Li, J., Chen, G. B., Rasheed, A., Li, D., Sonder, K., Zavala Espinosa, C., ... & Li, H. (2019). Identifying loci with breeding potential across temperate and tropical adaptation via EigenGWAS and EnvGWAS. *Molecular ecology*, *28*(15), 3544-3560.

Li, Y., Zhu, J., Wu, H., Liu, C., Huang, C., Lan, J., ... & Xie, C. (2020). Precise base editing of non-allelic acetolactate synthase genes confers sulfonylurea herbicide resistance in maize. *The Crop Journal*, *8*(3), 449-456.

Li, J., Li, D., Espinosa, C. Z., Pastor, V. T., Rasheed, A., Rojas, N. P., ... & Li, H. (2021). Genome-wide analyses reveal footprints of divergent selection and popping-related traits in CIMMYT's maize inbred lines. *Journal of experimental botany*, *72*(4), 1307-1320.

Linck, E., & Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources*, *19*(3), 639-647.

Liu, X., Xu, X., Li, B., Wang, X., Wang, G., & Li, M. (2015). RNA-seq transcriptome analysis of maize inbred carrying nicosulfuron-tolerant and nicosulfuron-susceptible alleles. *International Journal of Molecular Sciences*, *16*(3), 5975-5989.

Loux, M. M., Doohan, D., Dobbels, A. F., Johnson, W. G., Young, B. G., Legleiter, T. R., & Hager, A. (2017). Weed control guide for Ohio, Indiana and Illinois.

Lu, F., Romay, M. C., Glaubitz, J. C., Bradbury, P. J., Elshire, R. J., Wang, T., ... & Buckler, E. S. (2015). High-resolution genetic mapping of maize pan-genome sequence anchors. *Nature communications*, *6*(1), 1-8.

Matz, S. A. (1984). Snacks based on popcorn. In *Snack food technology* (pp. 138-149). Springer, Dordrecht.

Matz, Samuel A. "Snacks based on popcorn." *Snack food technology*. Springer, Dordrecht, 1984. 138-149.

Mazur, B. J., Chui, C. F., & Smith, J. K. (1987). Isolation and characterization of plant genes coding for acetolactate synthase, the target enzyme for two classes of herbicides. *Plant Physiology*, *85*(4), 1110-1117.

Merrill, W. L., Hard, R. J., Mabry, J. B., Fritz, G. J., Adams, K. R., Roney, J. R., & MacWilliams, A. C. (2009). The diffusion of maize to the southwestern United States and its impact. *Proceedings of the National Academy of Sciences*, *106*(50), 21019-21026.

Miller, J. M., Cullingham, C. I., & Peery, R. M. (2020). The influence of a priori grouping on inference of genetic clusters: simulation study and literature review of the DAPC method. *Heredity*, *125*(5), 269-280.

Nordby, J. N., Williams, M. M., Pataky, J. K., Riechers, D. E., & Lutz, J. D. (2008). A common genetic basis in sweet corn inbred Cr1 for cross sensitivity to multiple cytochrome P450-metabolized herbicides. *Weed Science*, *56*(3), 376-382.

Okonechnikov, K., Conesa, A., & García-Alcalde, F. (2016). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, *32*(2), 292-294.

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*(2), 289-290.

Peiffer, J. A., Romay, M. C., Gore, M. A., Flint-Garcia, S. A., Zhang, Z., Millard, M. J., ... & Buckler, E. S. (2014). The genetic architecture of maize height. *Genetics*, *196*(4), 1337-1356.

Prall, W., Hendy, O., & Thornton, L. E. (2016). Utility of a phylogenetic perspective in structural analysis of CYP72A enzymes from flowering plants. *PloS one*, *11*(9), e0163024.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945-959.

Reif, J. C., Hamrit, S., Heckenberger, M., Schipprack, W., Peter Maurer, H., Bohn, M., & Melchinger, A. E. (2005). Genetic structure and diversity of European flint maize populations determined with SSR analyses of individuals and bulks. *Theoretical and Applied Genetics*, *111*, 906-913.

Riechers, D. E., Kreuz, K., & Zhang, Q. (2010). Detoxification without intoxication: herbicide safeners activate plant defense gene expression. *Plant Physiology*, *153*(1), 3-13.

Romay, M. C., Millard, M. J., Glaubitz, J. C., Peiffer, J. A., Swarts, K. L., Casstevens, T. M., ... & Gardner, C. A. (2013). Comprehensive genotyping of the USA national maize inbred seed bank. *Genome biology*, *14*(6), 1-18.

Saika, H., Horita, J., Taguchi-Shiobara, F., Nonaka, S., Nishizawa-Yokoi, A., Iwakami, S., ... & Toki, S. (2014). A novel rice cytochrome P450 gene, CYP72A31, confers tolerance to acetolactate synthase-inhibiting herbicides in rice and Arabidopsis. *Plant physiology*, *166*(3), 1232-1240.

Santacruz-Varela, A., Widrlechner, M. P., Ziegler, K. E., Salvador, R. J., Millard, M. J., & Bretting, P. K. (2004). Phylogenetic relationships among North American popcorns and their evolutionary links to Mexican and South American popcorns. *Crop Science*, *44*(4), 1456-1467.

Serna-Saldivar, S. O. (2022). Popcorn and Other Puffed Grains. In *Snack Foods* (pp. 201-220). CRC Press.

Senhorinho, H. J. C., Coan, M. M. D., Marino, T. P., Kuki, M. C., Pinto, R. J. B., Scapim, C. A., & Holland, J. B. (2019). Genomic-wide association study of popping expansion in tropical popcorn and field corn germplasm. *Crop Science*, *59*(5), 2007-2019.

Smith, A. F. (1999). *Popped culture: A social history of popcorn in America*. Univ of South Carolina Press.

Swarts, K., Li, H., Romero Navarro, J. A., An, D., Romay, M. C., Hearne, S., ... & Bradbury, P. J. (2014). Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. *The Plant Genome*, *7*(3), plantgenome2014-05.

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., & Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*, *20*(8), 467-484.

Team, R. C. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2012.

Thyssen, G. N., Naoumkina, M., McCarty, J. C., Jenkins, J. N., Florane, C., Li, P., & Fang, D. D. (2018). The P450 gene CYP749A16 is required for tolerance to the sulfonylurea herbicide trifloxysulfuron sodium in cotton (Gossypium hirsutum L.). *BMC plant biology*, *18*(1), 1-8.

Turner, S. D. (2018). qqman: an R package for visualizing GWAS results using QQ and manhattan plots. J. Open Source Softw. 3: 731.

USDA NASS. (2022). Quick Stats. US Department of Agriculture. National Agricultural Statistics Service. https://quickstats.nass.usda.gov/

Van Heerwaarden, J., Doebley, J., Briggs, W. H., Glaubitz, J. C., Goodman, M. M., de Jesus Sanchez Gonzalez, J., & Ross-Ibarra, J. (2011). Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proceedings of the National Academy of Sciences*, *108*(3), 1088-1092.

Vázquez-Carrillo, M. G., Santiago-Ramos, D., & de Dios Figueroa-Cárdenas, J. (2019). Kernel properties and popping potential of Chapalote, a Mexican ancient native maize. *Journal of Cereal Science*, *86*, 69-76.

Vigouroux, Y., Glaubitz, J. C., Matsuoka, Y., Goodman, M. M., Sánchez G, J., & Doebley, J. (2008). Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. *American journal of botany*, *95*(10), 1240-1253.

Vittorazzi, C., Junior, A. A., Guimarães, A. G., Silva, F. H. L., Pena, G. F., Daher, R. F., ... & Lima, V. J. (2018). Evaluation of genetic variability to form heterotic groups in popcorn. *Genet. Mol. Res*, *17*, 18083.

Wallace, J. G., Larsson, S. J., & Buckler, E. S. (2014). Entering the second century of maize quantitative genetics. *Heredity*, *112*(1), 30-38.

Wang, L., Beissinger, T. M., Lorant, A., Ross-Ibarra, C., Ross-Ibarra, J., & Hufford, M. B. (2017). The interplay of demography and selection during maize domestication and expansion. *Genome biology*, *18*(1), 1-13.

White, M. R., Mikel, M. A., de Leon, N., & Kaeppler, S. M. (2020). Diversity and heterotic patterns in North American proprietary dent maize germplasm. *Crop Science*, *60*(1), 100-114.

Wickham, H., Chang, W., & Wickham, M. H. (2016). Package 'ggplot2'. *Create elegant data visualisations using the grammar of graphics. Version*, *2*(1), 1-189.

Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., ... & Liu, X. (2021). rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics, proteomics & bioinformatics*, *19*(4), 619-628.

Yu, D., Wang, H., Gu, W., Qin, T., Sun, P., Lu, Y., ... & Zheng, H. (2021). Genetic diversity and population structure of popcorn germplasm resources using genome-wide SNPs through genotyping-by-sequencing. *Genetic Resources and Crop Evolution*, *68*(6), 2379-2389.

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. Y. (2017). ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*(1), 28-36.

Yuan, J. S., Tranel, P. J., & Stewart Jr, C. N. (2007). Non-target-site herbicide resistance: a family business. *Trends in plant science*, *12*(1), 6-13.

Ziegler, Kenneth E. "Popcorn." *Specialty corns*. CRC Press, 2000. 211-246.

Figure 1: Identity by state (IBS) distribution across 320 (A) heterogeneous and (B) inbred popcorn accessions.

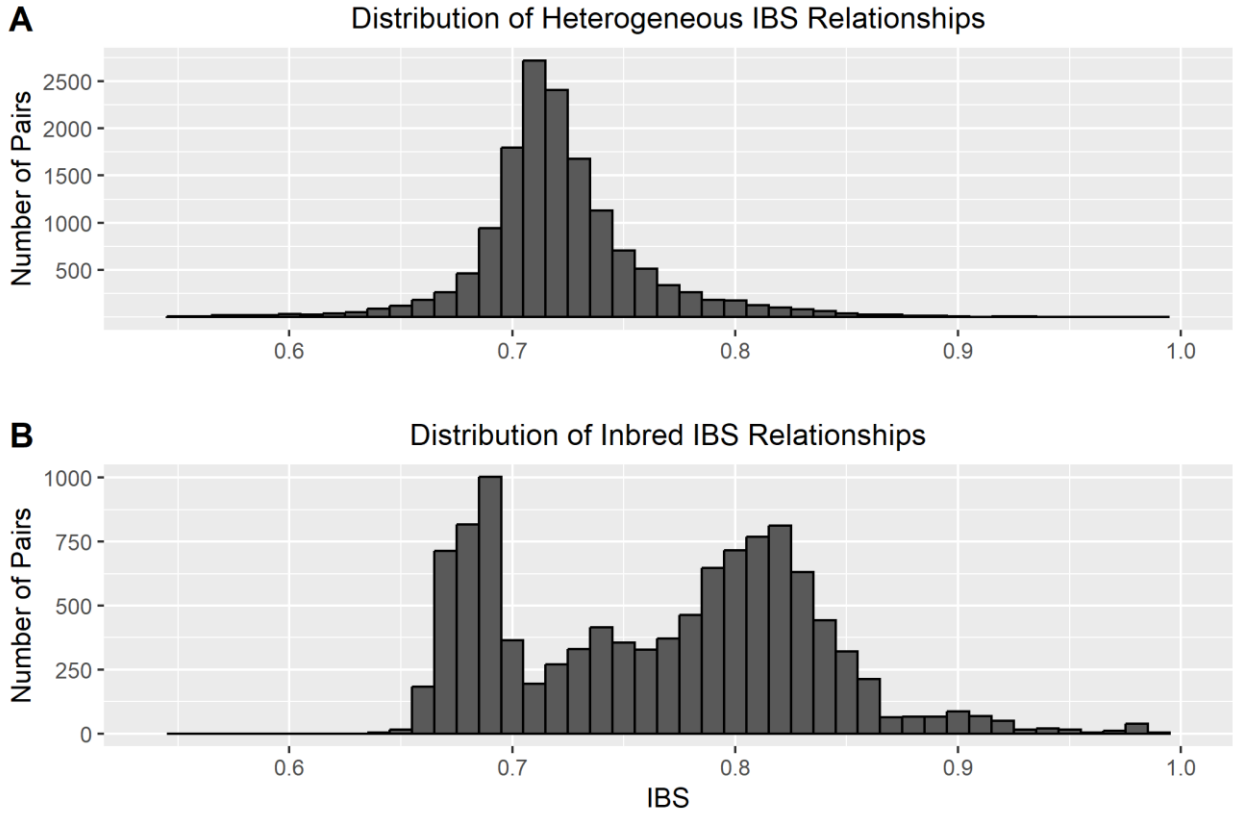

Figure 2: Bar plot of assignment values from ADMIXTURE analysis. Each bar corresponds to one of 320 (A) heterogeneous, (B) inbred, or (C) combined popcorn accessions, and is ordered by the likelihood of assignment. The complete list of assignments can be found in Supplementary Tables S2-S4.

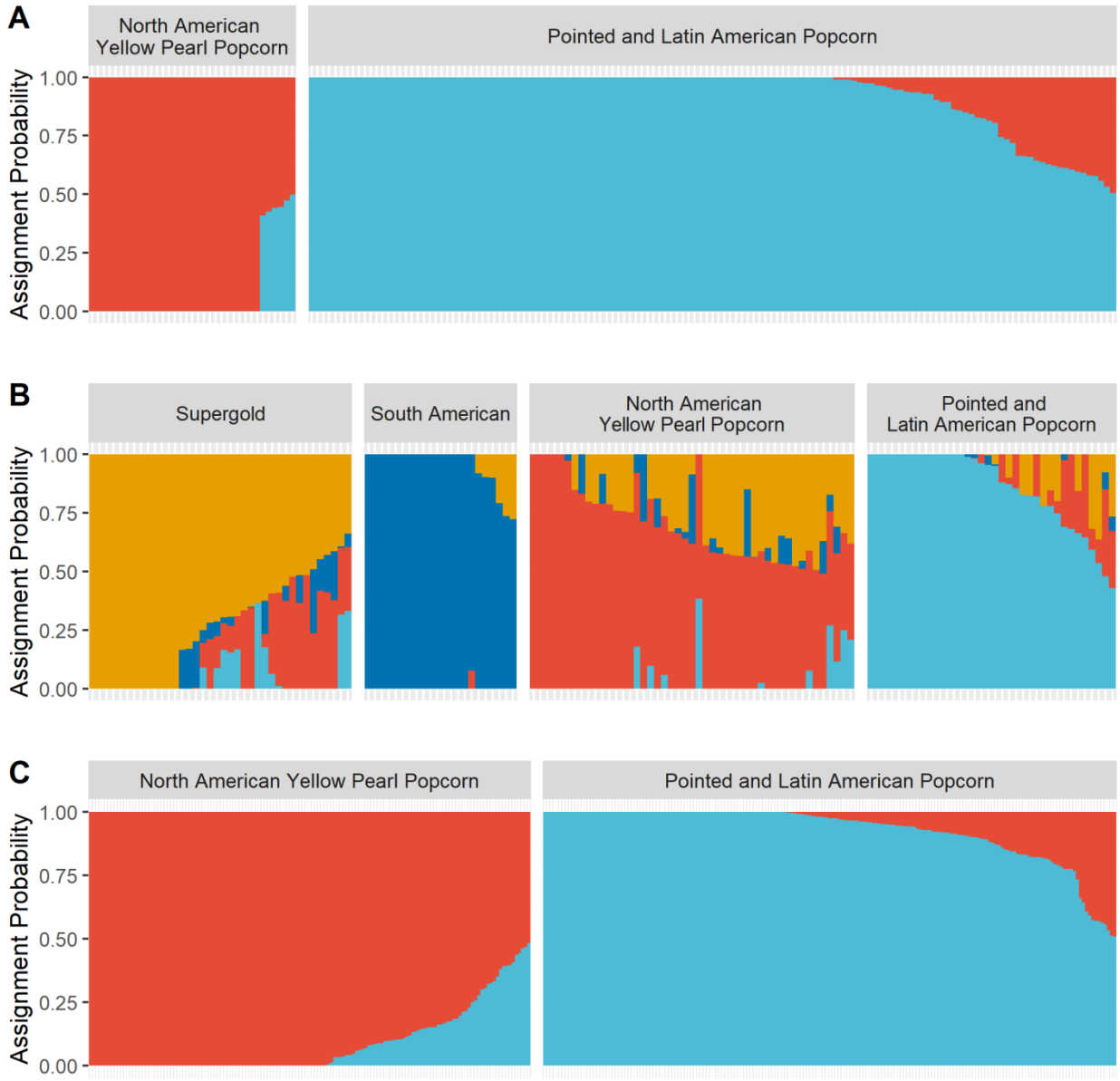

Figure 3: Principal coordinate analysis (PCoA) of 320 (A) heterogeneous, (B) inbred, and (C) combined popcorn accessions estimated with the Core Analysis marker set and associated scree plot

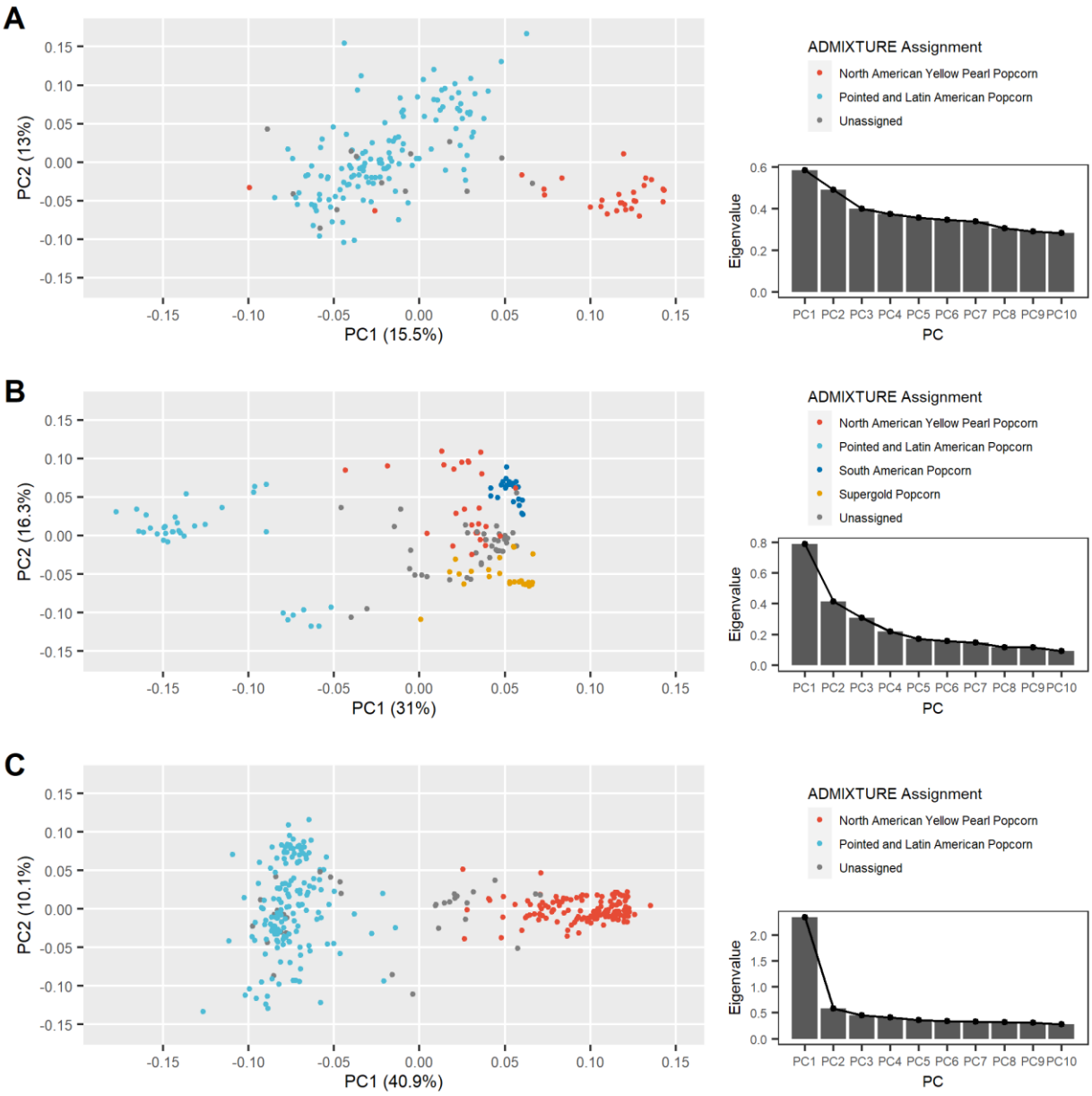

Figure 4: Neighbor-joining tree of 320 popcorn accessions and 4 dent accessions based on a modified Euclidean distance matrix calculated from the Core Analysis marker set. The three heterotic pools of North American Yellow Pearl Popcorns are labeled, along with the dent outgroup. Accessions SONO 147 and BRAZ 2832 are marked with an asterisk (∗). Accessions belonging to Curagua are marked with a dot (•).

Figure 5: Minor allele frequency (MAF) distributions of (A) heterogeneous and (B), inbred groups of popcorn. Y-axis shows the proportion of SNPs, as differing counts of monomorphic SNPs were removed from each population prior to calculating frequencies

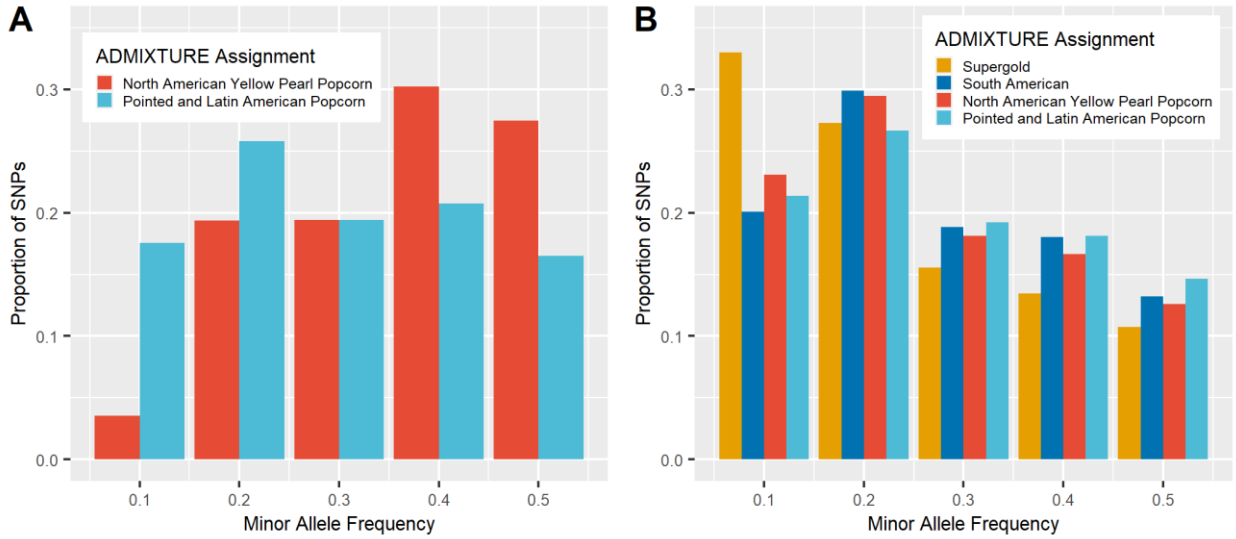

Figure 6: Median linkage disequilibrium (LD) decay of (A) heterogeneous and (B), inbred groups of popcorn. Pairwise comparisons were binned by distance using logarithmically increasing bins

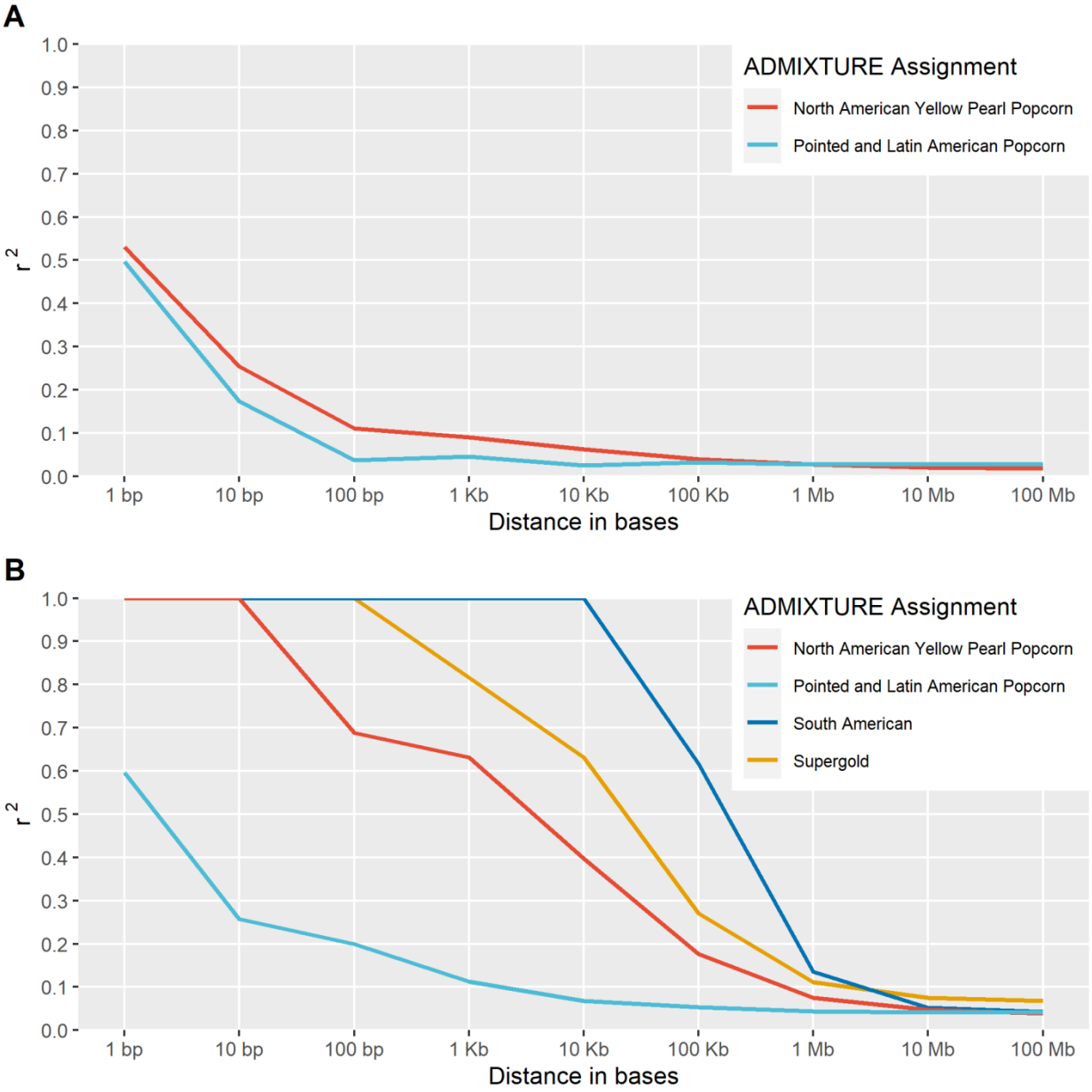

Figure 7: Manhattan plot of genome-wide association study (GWAS) of kernel color from the Filtered GWAS set. The red line represents the threshold after correcting for multiple testing using a Bonferroni correction

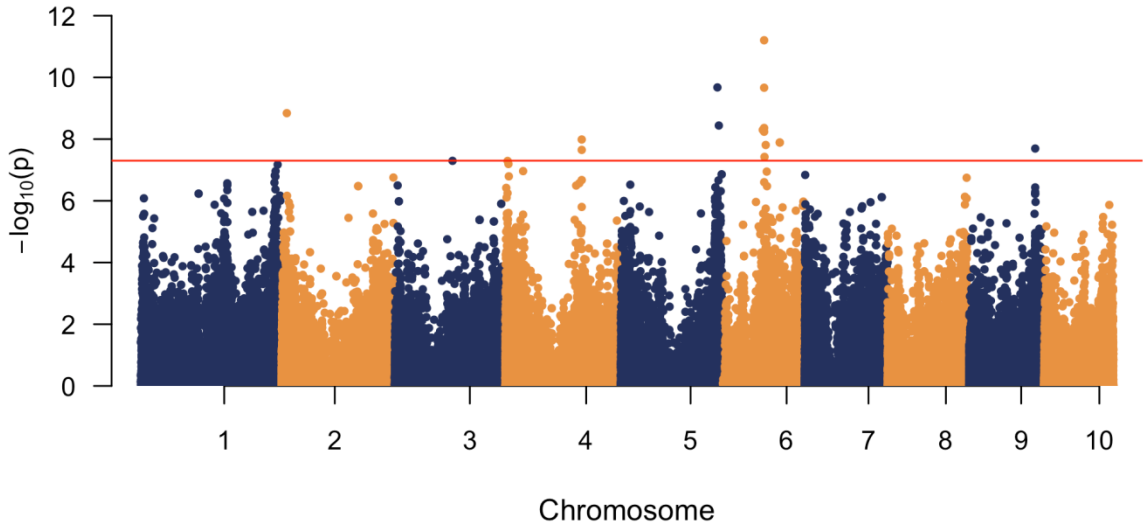

Figure 8: Manhattan plot of genome-wide association study (GWAS) of nicosulfuron injury from the Filtered GWAS set. The red line represents the threshold after correcting for multiple testing using a Bonferroni correction