Molecular Ecology

WILEY

MOLECULAR ECOLOGY

ORIGINAL ARTICLE OPEN ACCESS

Structural Variants Contribute to Phenotypic Variation in Maize

Nathan S. Catlin[1,2,3] | Husain I. Agha[1,3] | Adrian E. Platts[1] | Manisha Munasinghe[4] | Candice N. Hirsch[5] | Emily B. Josephs[1,2,3]

[1]Department of Plant Biology, Michigan State University, East Lansing, Michigan, USA | [2]Ecology, Evolution, and Behavior Program, Michigan State University, East Lansing, Michigan, USA | [3]Plant Resilience Institute, Michigan State University, East Lansing, Michigan, USA | [4]Department of Plant and Microbial Biology, University of Minnesota, St. Paul, Minnesota, USA | [5]Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, Minnesota, USA

Correspondence: Nathan S. Catlin (catlinna@msu.edu) | Emily B. Josephs (josep993@msu.edu)

Received: 1 July 2024 | **Revised:** 4 December 2024 | **Accepted:** 31 December 2024

Handling Editor: Sarah Turner-Hissong

Funding: This work was supported by the National Science Foundation IOS-1934384 to C.N.H and E.B.J. and a Postdoctoral Research Fellowship in Biology under Grant No. IOS-2010908 to M.M., National Institutes of Health R35-GM142829 to E.B.J, and USDA NIFA Project MICL02656 to E.B.J.

Keywords: agriculture | life history evolution | phenotypic plasticity | population genetics – empirical | quantitative genetics | structural variants

ABSTRACT

Comprehensively identifying the loci shaping trait variation has been challenging, in part because standard approaches often miss many types of genetic variants. Structural variants (SVs), especially transposable elements (TEs), are likely to affect phenotypic variation but we lack methods that can detect polymorphic SVs and TEs using short-read sequencing data. Here, we used a whole genome alignment between two maize genotypes to identify polymorphic SVs and then genotyped a large maize diversity panel for these variants using short-read sequencing data. After characterising SV variation in the panel, we identified SV polymorphisms that are associated with life history traits and genotype-by-environment (GxE) interactions. While most of the SVs associated with traits contained TEs, only two of the SVs had boundaries that clearly matched TE breakpoints indicative of a TE insertion, while the other polymorphisms were likely caused by deletions. One of the SVs that appeared to be caused by a TE insertion had the most associations with gene expression compared to other trait-associated SVs. All of the SVs associated with traits were in linkage disequilibrium with nearby single nucleotide polymorphisms (SNPs), suggesting that the approach used here did not identify unique associations that would have been missed in a SNP association study. Overall, we have (1) created a technique to genotype SV polymorphisms across a large diversity panel using support from genomic short-read sequencing alignments and (2) connected this presence/absence SV variation to diverse traits and GxE interactions.

1 | Introduction

A central question of evolutionary biology is how different types of mutations—single nucleotide polymorphisms (SNPs), insertion–deletion polymorphisms, copy number variants, translocations and transposable element (TE) insertions—shape the phenotypic diversity observed in nature (Mitchell-Olds, Willis, and Goldstein 2007). Much recent effort has focused on characterising structural variants (SVs): Tens of thousands of SVs have been identified in plant genomes (Darracq et al. 2018; Yang et al. 2019; Schatz 2018; Alonge et al. 2020; Zhou et al. 2022; Qin et al. 2021; Hämälä et al. 2021) and specific SVs have been shown to affect important phenotypic traits in plants, including climate resilience in *Arabidopsis thaliana*, disease resistance and domestication traits in maize and rice, and frost tolerance in wheat (Beló et al. 2010; Cao et al. 2011; Sieber et al. 2016;

Springer et al. 2009; Xu et al. 2012). In addition, maize SVs are predicted to be up to 18-fold enriched for alleles affecting phenotypes when compared to SNPs (Chia et al. 2012). These findings suggest that characterising SV variation will be a crucial part of mapping genotypes to phenotypes.

A subset of SVs, TEs, are particularly interesting potential contributors to phenotypic variation (Lisch 2013; Catlin and Josephs 2022). TE content and polymorphism are shaped by a complex interplay of selection at the TE and organismal level (Charlesworth and Charlesworth 1983; Ågren and Wright 2011) and there are many examples of TE variation affecting phenotypes (Hirsch and Springer 2017; Lisch 2013). For example, a TE insertion in the regulatory region of the *teosinte branched1* (Zm*TB1*) gene in maize enhances gene expression, causing the upright branching architecture in maize compared to its progenitor, teosinte (Studer et al. 2011). TE insertions also affect flesh colour in grapes and fruit colour and shape in tomato (Fray and Grierson 1993; Kobayashi, Goto-Yamamoto, and Hirochika 2004; Knaap et al. 2004; Shimazaki et al. 2011; Domnguez et al. 2020). These phenotypic effects may result from changes in gene expression: TE activation can disrupt or promote gene expression (Hirsch and Springer 2017; Fueyo et al. 2022), and the industrial melanism phenotype in British peppered moths, *Biston betularia*, results from TE-induced overexpression of a gene responsible for pigment production (Hof et al. 2016). TEs often activate (i.e., express and/or mobilise) in response to stress in many eukaryotes, including maize (Makarevitch et al. 2015; Liang et al. 2021), *Arabidopsis* (Wang et al. 2022; Sun et al. 2020) and *Drosophila melanogaster* (de Oliveira et al. 2021; Milyaeva et al. 2023), suggesting that they may contribute to trait variation in stressful environments. However, we lack systematic studies of how TEs in general affect phenotypic variation or how TEs may contribute to genotype-by-environment interactions outside of the context of stress.

Characterising genomic variation for SVs and TEs has been challenging, especially in highly repetitive plant genomes where it is often difficult to uniquely align short-reads to a reference genome. Recent studies have shown that attempts to assemble SVs solely with short-read sequencing data can greatly underestimate the total number of SVs present in a population (Huddleston et al. 2017; Audano et al. 2019; Cameron, Di Stefano, and Papenfuss 2019; Ebert et al. 2021). Some estimates for the accuracy of SV discovery with short-read sequencing are as low as 11% in humans because of the inability of short-reads to align within highly repetitive regions, span large insertions or concordantly align across SV boundaries (Lucas Lledó and Cáceres 2013). However, short-read sequencing from a population of grapevine cultivars has been used to genotype SVs by ascertaining SV polymorphisms between two reference genomes and calling these SVs within the population (Zhou et al. 2019).

The increasing availability of long-read sequencing has opened up an opportunity to identify SVs that would have been missed using short-read data. For example, long reads have been used to identify SVs associated with traits in a set of 100 tomato accessions (Alonge et al. 2020). In other systems without enough long-read sequenced genotypes to directly look for associations between SVs and phenotype, researchers have started with SVs detected in a smaller subset of individuals with reference assemblies and then genotyped in a larger mapping panel of individuals with short-read sequencing data. Researchers have used pan-genome graph methods to identify SVs in a smaller number of reference sequences and then genotyped in a larger sample of short-read sequenced lines in *Arabidopsis thaliana* (Kang et al. 2023), soybean (Liu et al. 2020), rice (Qin et al. 2021) and tomato (Zhou et al. 2022). These studies have confirmed that SVs are important for trait heritability (Zhou et al. 2022). However, graph genome approaches are challenging for plants with large genomes and have not yet been widely adopted. For example, a haplotype graph has been generated for 27 maize inbred lines, but not for a wider diversity panel (Franco et al. 2020). Additionally, work using short-read alignments and pan-genome approaches have identified SVs in maize and found that SVs contributed to trait heritability (Gui et al. 2022). Approximately 60% of these SVs were related to TEs but no clear links between SV polymorphisms and TE insertions were made (Gui et al. 2022). Plants with large genomes are not only important for a number of practical reasons, but they also may evolve differently than plant species with small genomes because they have different genetic architectures underlying trait variation (Mei et al. 2018). Understanding how SVs and TEs contribute to trait variation in plants with large genomes is key for comprehensively understanding the importance of these variants in general.

To address the gap in understanding how SVs and TEs contribute to trait variation in a species with a large genome, we identified SVs found from the alignment of two maize reference assemblies using short-reads that overlap the SV junctions. This type of approach has been used previously in a few other systems (Wang et al. 2020; Zhou et al. 2019). Here, we investigated the relationship between SV variation and phenotype in a diverse set of maize inbred lines in the Buckler–Goodman association panel (Flint-Garcia et al. 2005). After identifying SVs that differ between two accessions, B73 and Oh43, we genotyped 277 maize lines present in a larger mapping panel for the SV alleles. We detected SV polymorphisms that varied across the panel and linked these polymorphisms to phenotypic variation, GxE and gene expression.

2 | Materials and Methods

2.1 | Structural Variant Identification

An ascertainment set of SVs that differ between B73 and Oh43 was identified by Munasinghe et al. (2023). The methods used by Munasinghe et al. (2023) were not able to identify inversions. These genotypes were chosen to call SV presence/absence because they are both in the Buckler–Goodman association panel but come from different germplasm pools (Gage et al. 2019). Ascertainment set SVs were filtered to only contain those that had 300 bps of colinear sequence determined by AnchorWave (Song et al. 2022) in the immediate upstream and downstream regions flanking SV junctions. The apparent insertion and 300 bp flanking region on either side were extracted to create a FASTA file containing SV-present alleles. The corresponding site in the other genome where the SV was absent and 300 bp flanking sequences were also extracted and combined in the final FASTA file to serve as the SV-absent allele sequence. Ultimately, this

FASTA file was used as a set of pseudoreference alleles to call SV polymorphism in individuals with only short-read sequence data (Figure S1).

2.2 | SV Presence/Absence Genotyping

To call presence or absence for each SV, we aligned genomic short-read data for 277 inbred maize genotypes from the Buckler–Goodman association panel to the generated FASTA files with SV present and absent alleles (Flint-Garcia et al. 2005; Bukowski et al. 2018). Illumina adapters and low quality sequences were removed using Trimmomatic v0.39 (Bolger, Lohse, and Usadel 2014). PCR duplicate reads were also filtered out using the -r option within the *markdup* function in SAMtools v1.15.1 (Danecek et al. 2021). Surviving paired-end reads were merged into a master FASTQ file for each genotype and aligned to pseudoreference alleles using HISAT2 (Sirén, Välimäki, and Mäkinen 2014). The aligned dataset was filtered to only contain concordant, uniquely mapping reads. We used read-depth for each upstream and downstream SV boundary to support the presence or absence of SVs (Figure 1). Read coverage at each SV boundary was calculated using the *coverage* function within bedtools v2.30.0 (Quinlan and Hall 2010).

First, we filtered out SVs where we were unable to use short-read data from B73 and Oh43 to correctly identify their respective SV genotypes. In these cases, short-read data mapped better to the alternative genotype's allele than their own allele. For an SV within our ascertainment set to be retained for downstream genotyping in the Buckler–Goodman association panel, we required that: (1) upstream and downstream SV junctions had the same or higher read coverage from the genotype with the SV

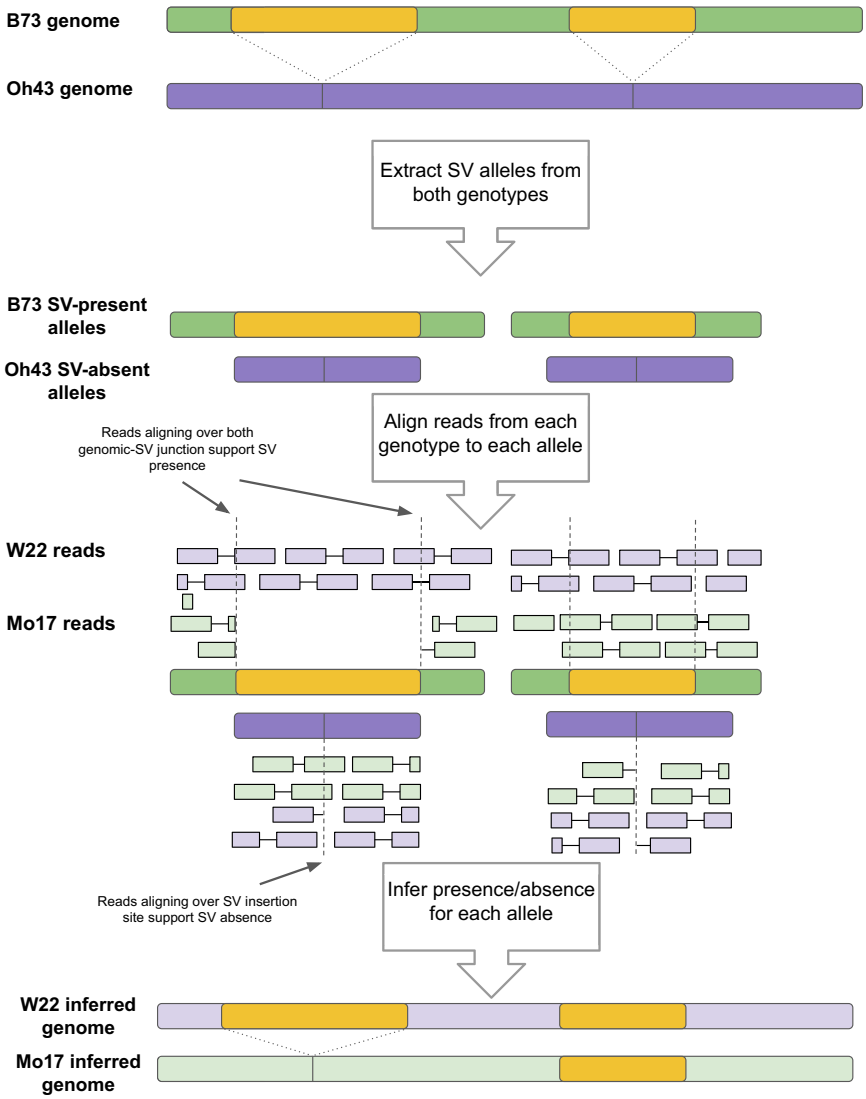

FIGURE 1 | Method to call SV presence/absence with short read genomic data—Using B73 and Oh43 as our ascertainment set, we first find polymorphic SVs between these two genotypes. To significantly improve read-mapping runtimes, we extract SVs and adjacent genomic sequences where SVs are present, while extracting only adjacent genomic regions at the polymorphic site where the SV is absent in the opposite genotype—termed pseudoreference SV alleles. Next, reads from a genotype of interest are mapped to these generated sequences. SVs can then be inferred as present or absent based on their alignment to either allele.

than the other genotype and (2) no reads from the SV-present genotype spanned the insertion site for the genotype without the SV (Figure S2).

For the rest of the genotypes in the Buckler–Goodman association panel, SV-presence was supported in the query genotype if there was at least one read spanning the upstream or downstream SV junction and there was no read coverage at the SV polymorphic site for the alternative SV-absent allele. An SV-absent allele is supported if at least one read spans across the SV polymorphic site but no reads map to either SV junction of the corresponding SV-present allele. SVs are ambiguous if reads from the query genotype map to both the SV-present allele junctions and the SV-absent insertion site.

2.3 | Calculating Linkage Disequilibrium Between SNPs and SVs

SNPs in variant call format (VCF) were collected from the third generation maize haplotype map (HapMap3) version 3.2.1 and coordinates were converted to the B73 NAM reference positions (version 5) using liftOverVCF in Picard tools (Pic 2019; Qiu et al. 2021). We removed SNPs with >10% missing data, a minor allele frequency (MAF) <10%, and those within SV regions, resulting in 16,435,136 SNPs in the final filtered dataset. Additionally, we appended polymorphic SV calls for each genotype in the HapMap3 dataset to the final VCF file. Because SV-present alleles were characterised for both B73 and Oh43, we used the start of the SV coordinate for SV-present alleles within B73 and the B73 insertion site for SVs present in Oh43 as the coordinate for LD analysis. Following methods from Qiu et al. (2021), we calculated LD between SNPs and nearby polymorphic SVs, excluding SNPs inside of SVs, using PLINK v1.9 (Chang et al. 2015), www.cog-genomics.org/plink/1.9/ with the following parameters: -make-founders, -r2 gz dprime with-freqs, -ld-window-r2 0, -ld-window 1,000,000, -ld-window-kb 1000 and -allow-extra-chr. We conducted a χ^2 test for differences in allele frequency between populations using the chisq.test() function in R (R Core Team 2024).

2.4 | Association Mapping

Polymorphic SVs across all query genotypes were converted to BIMBAM mean genotype format (Servin and Stephens 2007). SV-present alleles that were characterised as ambiguous were denoted as NA. We performed a genome-wide association study (GWAS) of SV presence/absence variants (PAVs) using phenotypes from Peiffer et al. (2014) and a linear mixed model (LMM) in GEMMA v0.98.03 (Zhou and Stephens 2012). The traits tested are best linear unbiased predictions (BLUPs) of the following: growing degree days to silking, growing degree days to anthesis, anthesis-silking interval measured in growing degree days, days to silking, days to anthesis, anthesis-silking interval measured in days, plant height, ear height, difference of plant height and ear height, ratio of ear height and plant height, and ratio of plant height and days to anthesis. To account for missing genotypic data for each SV, we required at least 90% of the genotypes to have presence/absence calls for relatedness matrix calculations and subsequent associations. All plots with genomic locations are shown with B73 coordinates, and Oh43 SV-present alleles were converted to B73 coordinates for display. To account for multiple-testing, we calculated a false discovery rate (FDR)-adjusted significance threshold (Benjamini and Hochberg 1995) to maintain an overall $\alpha = 5\%$ significance. Filtered SNPs from the HapMap3 dataset were also subjected to GWAS using the same methods as our polymorphic SV dataset.

In addition to the association analyses for main effects, we examined these data for genotype-by-environment interactions (GxE). For the 11 traits above, we used simple linear regression following the form of Finlay–Wilkinson (FW) regression (Finlay and Wilkinson 1963) to record the slope (i.e., reaction norm) and mean squared error (MSE) for each genotype using the linear model (lm) function in R;

$$y_{ij} = \beta_0 + \beta_1 x_j + \epsilon_{ij},$$

where β_0 and β_1 are the intercept and slope estimates for the i^{th} line, respectively, x_j is the average performance of all lines in the j^{th} environment and ϵ_{ij} is a random error term. We removed any lines which were not represented in at least six environments on a per trait basis to reduce the error in our estimates. This filtering resulted in a different number of individuals and markers used in each FW model (ranging from 245 to 274 individuals per trait). We then performed GWAS of SV PAVs using slope and MSE estimates for each trait as quantitative phenotypes in GEMMA as before.

2.5 | Gene Expression

We used previously collected gene expression data for ~37,000 maize genes (Kremling et al. 2018) to test for differential gene expression between SV genotypes at the loci identified in the association mapping analyses. We compared expression between SV genotypes for three tissue types: the tip of germinating shoots, the base of the third leaf and the tip of the third leaf. Library sizes were normalised using DESeq2 (Love, Anders, and Huber 2014) and we filtered the gene set to contain only genes with expression in 70% of individuals above 10 reads per the median library size (approx 0.5 counts per million) using the edgeR package in R (Robinson, McCarthy, and Smyth 2010), resulting in an average of 12,703 genes per SV identified in the GWAS. Finally, we used edgeR to test for differential expression by first building generalised linear models to model expression between genotypes and then testing for significance using the *F*-test. *p*-values were adjusted using FDR to maintain an overall significance threshold of $\alpha = 5\%$.

3 | Results

3.1 | Polymorphic SVs in the Diversity Panel

We genotyped SV polymorphisms for 277 maize genotypes at SVs segregating between B73 and Oh43 by aligning short reads from the genotypes to each SV allele and counting reads spanning genomic-SV junctions and SV polymorphic sites. Out of 98,422 polymorphic SVs between B73 and Oh43, we filtered out SVs where short reads from B73 and Oh43 did not clearly align

to the correct allele. After this filtering step, we were able to determine the genotype of 64,956 SVs in the Buckler–Goodman association panel (Figure S2). The largest proportion of these SVs were those classified as 'TE = SV' (21,103, 32.5%), followed by 'multi TE SVs' (18,326, 28.2%), 'incomplete TE SVs' (10,928, 16.8%), 'no TE SVs' (8842, 13.6%) and 'TE within SVs' (5757, 8.9%) (Figures S3 and S4). The proportions of SVs for each category are consistent with those prior to filtering. For more information about how SVs are classified into TE groupings, see Munasinghe et al. (2023).

For subsequent analyses, we filtered the SV dataset to only include variants with a MAF ≥ 10% and presence/absence calls for at least 90% of genotypes, resulting in the retention of 3,087 SV alleles (4.75% of dataset) (Figure S5). Filtering on missing data and MAF removed many SVs because many individuals in the dataset have low realised sequencing coverage when mapped to the B73 reference assembly. There is a median coverage of 2.68, ranging from 0.031 in the A554 genotype to 19.47 in B57. Read depth per individual was negatively correlated with percent missing SV data per individual ($p = 2.4 \times 10^{-5}$) (Figures S6 and S7), suggesting that missing data for SVs results from not having enough reads covering the junction sites. This pattern suggests that this method needs a minimum average read depth of 5 to successfully genotype SVs at most sites, although this number will likely vary by species.

We investigated the frequency spectrum of SV polymorphisms in the Buckler–Goodman association panel by calculating the frequency of the allele with a putative insertion (or lacking a putative deletion). As these SVs were initially identified as being polymorphic between two individuals, it was not surprising to see that many of the SVs were at moderate frequency in the population (Figures 2 and S3). For most SVs, the SV-present allele was more common than the SV-absent allele. This pattern is consistent with the polymorphism being caused by a deletion and the longer 'insertion' allele being the ancestral type, and so present at higher allele frequencies in the population. The frequency spectrum was relatively consistent across SV types (Munasinghe et al. 2023).

3.2 | SV Genotypes Are Associated With Phenotypic Traits

In the GWAS, SV presence/absence was significantly associated (FDR < 0.05) with four out of the 11 traits tested: growing degree days to anthesis, days to silking, days to anthesis and ear height (Figures 3 and S8). All four SVs associated with traits contained TE sequences but none had boundaries that matched TE boundaries ('TE = SV'), suggesting that the polymorphisms were the result of deletions, not TE insertions (Figure 4).

The SV associated with growing degree days to anthesis is within B73 on chromosome seven, 54 bp upstream of the B73 gene Zm00001eb330210 (syntenic with Oh43 gene Zm00039ab336990) (Figures 3 and 4A). There are no currently known functions for these genes in maize, nor their orthologues in other species including sorghum, foxtail millet, rice or *Brachypodium distachyon*. There is evidence of increased

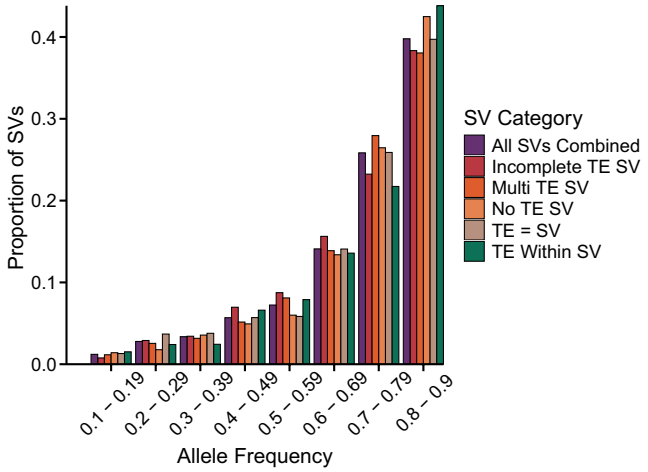

FIGURE 2 | Site-frequency Spectrum of SVs—SVs were filtered to only contain those with a minor allele frequency ≥ 10% and ≤ 10% missing data ($n = 3087$). The SFS is unfolded and displays the frequency of the allele with the putative insertion (or that is lacking a deletion).

expression in these genes in maize in whole seed, endosperm, and embryo for most 2-day increments post pollination (Walley et al. 2016). This SV contained a mutator TE within it, but the SV boundaries did not match the TE boundaries.

One SV polymorphism was associated with both days to silking and days to anthesis. This SV is present on chromosome three in Oh43 and is a large, ~52 kb multi TE SV composed primarily of Ty3 elements (Figures 3 and 4B). This region is nearly 215 kb away from the nearest gene. An additional SV associated with days to silking is located on chromosome 10 and contains ~43.5 kb of multiple Ty3 TEs (Figures 3 and 4C). This SV, present in B73 and absent in Oh43, is 2091 bp upstream of the gene Zm00001eb411130 (syntenic with the Oh43 gene Zm00039ab420040). Zm00001eb411130, which is also called ZmMM1, is a MADS-box gene and is orthologues with the OsMADS13 gene in rice and the STK gene in *Arabidopsis thaliana*. OsMADS13's expression in rice is restricted to the ovule and controls both ovule identity and meristem determinancy during ovule development (Lopez-Dee et al. 1999; Dreni et al. 2007; Li et al. 2011). Similar to OsMADS13, STK in *Arabidopsis thaliana*, which encodes for a MADS-box transcription factor, is expressed in the early floral development in the ovule. Additionally, STK determines ovule identity and also regulates a network of genes that controls seed development and fruit growth (Mizzotti et al. 2014; Di Marzo et al. 2020). Both OsMADS13 and STK are members of the D-class genes in the ABCDE model for floral development.

The SV associated with ear height contains a partial sequence of a mutator DNA transposon and is on Oh43 chromosome four within an intron of gene Zm00039ab208360 (syntenic with B73 gene Zm00001eb203840) (Figures 3 and 4D). This gene, also called *traf42*, is a tumour receptor-associated factor (TRAF) and codes for a BTB/POZ domain-containing protein *POB1*. Although TRAF domain containing proteins are ubiquitous across eukaryotes, there are far more genes encoding TRAF domains in plants compared to animals (Oelmüller et al. 2005; Cosson et al. 2010). In maize, *traf42* mediates protein–protein interactions (Dong et al. 2017) and mutations in the maize gene ZmMAB1, which

contains a TRAF domain and is exclusively expressed in the germline cause chromosome segregation defects during meiosis (Juranić et al. 2012). Additionally, *POB1* is involved in drought tolerance in the Antarctic moss, *Sanionia uncinata* (Park et al. 2018).

3.3 | SV Genotypes Are Associated With GxE

We detected five significant associations (FDR < 0.05) between SV presence/absence and one of two measures of plasticity (FW regression slope and MSE) for four of the eleven traits tested: the ratio of plant height and days to anthesis (MSE), growing degree days to silking (MSE), days to silking (slope) and days to anthesis (slope) (Figures 3 and S9). Four of the five SVs identified contained TE sequence and two SVs appeared to be directly caused by TE insertions.

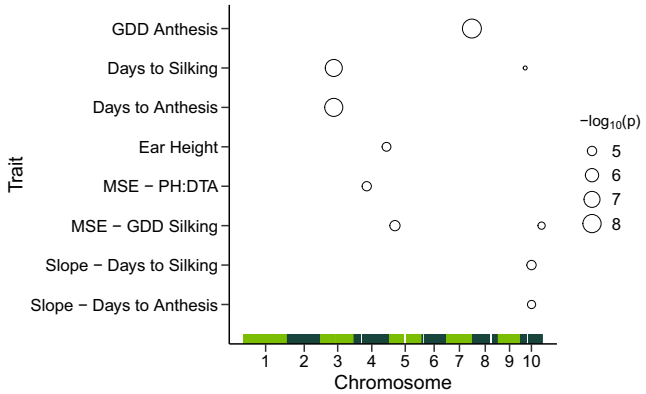

FIGURE 3 | Genomic positions and *p*-values for eight traits and nine markers with significant SV presence/absence associations—Bars at bottom represent the genomic positions for the 3087 SV markers used in the association panel, with chromosomes in alternating colours. Points are sized according to the $-\log_{10}(p)$ (GDD: growing degree days; MSE: mean squared error; PH:DTA: ratio of plant height to days to anthesis). Note that the same SV was associated with Days to Silking and Days to Anthesis so there are 10 points total.

On chromosome four, we detected an association between an SV and the MSE of the ratio of plant height to days to anthesis across growing locations. This SV appeared to be caused by a partial deletion of a Ty3-like LTR retrotransposon and was not proximal to any gene models in either the Oh43 or B73 alignments.

On chromosome five, we detected an association between an SV and the MSE of growing degree days to silking across growing locations. This SV appeared to be caused by a partial deletion of a hAT TIR transposon but was not proximal to any gene model in either the Oh43 or B73 alignments.

On chromosome 10, we detected three associations between SVs and plasticity: the slope of days to silking, the slope of days to anthesis, and the MSE of growing degree days to silking. The SV associated with days to silking appeared to be the direct result of an insertion of a PIF Harbinger TIR transposon, the SV associated with the MSE of growing degree days to silking appeared to be an insertion of a PIF Harbinger TIR transposon, but the SV associated with the slope of days to anthesis did not contain TE sequence. The SV associated with the slope of days to silking was 713 bp from the uncharacterized Oh43 gene Zm00039ab424300 (a syntelog of B73 gene Zm00001eb415280), while the SVs associated with the slope of days to anthesis and the MSE of growing degree days to silking were not proximal to any B73 or Oh43 gene model.

3.4 | SV Genotypes Are Associated With Differential Gene Expression

We tested for associations between the genotypes of the nine SVs identified by GWAS and gene expression data from three tissues and detected associations for 29 genes (Figure 5). Differentially expressed genes were not immediately proximal to the SVs they

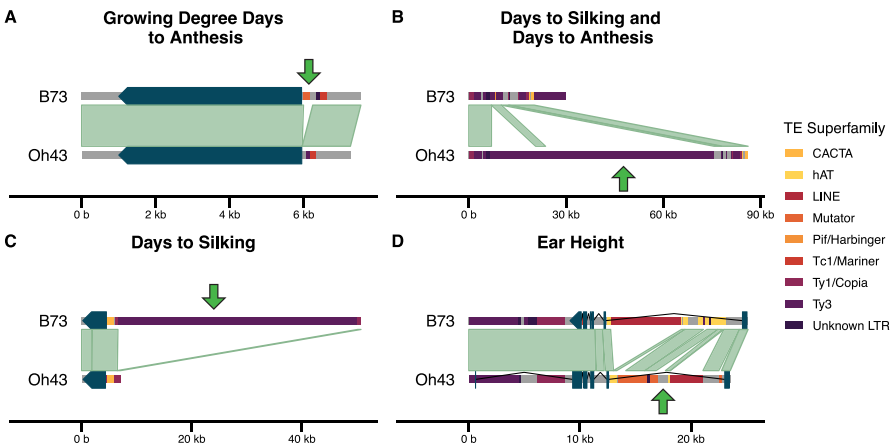

FIGURE 4 | Trait-associated structural variant polymorphisms between B73 and Oh43—Green arrows point to polymorphic SV regions that are associated with traits. Alignable regions are shown as green bars between genotypes. TEs are displayed inline and, therefore, do not display overlapping or nested TEs. (A) A mutator TE within an SV is present in B73 and absent in Oh43. This SV is 54 bp upstream of the B73 gene Zm00001eb330210, syntenic with Oh43 gene Zm00039ab336990. (B) A large SV containing multiple Ty3 TEs is present in Oh43 and absent in B73. This intergenic SV is approximately 215 kb from the nearest gene. (C) A multi TE SV composed entirely of Ty3 TEs is present in B73 and 2091 bp upstream of the gene Zm00001eb411130 (syntenic with Oh43 gene Zm00039ab420040). (D) A polymorphic incomplete TE SV is located within the Oh43 gene Zm00039ab208360 is present in Oh43 and absent in B73.

were associated with (the closest differentially expressed gene was 911 kb from the associated SV) and most were on different chromosomes. Of the 29 significantly associated genes, three genes present in the B73v3 reference alignment were not present in the B73v5 alignment and were removed from further consideration. Of the 26 remaining genes, 11 were associated with a single SV on chromosome 10 for the MSE of growing degree days to silking, which was coded as 'TE = SV'. The remaining six SVs were associated with between one and four differentially expressed genes and of those six SVs, three contained complete TE sequences, two contained incomplete TEs, and one did not contain any TE sequence. Of the three tissues tested, 16 genes were significantly differentially expressed solely in shoot tissue, seven in the tip of L3, two in the base of L3 and one was differentially expressed in both the shoot tissue and the base of L3.

3.5 | Most SVs Are in Linkage Disequilibrium With SNPs

All SV alleles used in the GWAS are within 1 Mb (mean distance of 649 bps) from the nearest SNP present in the HapMap3 dataset (Figure S10) and all SVs have an $r^2 > 0.1$ with at least one nearby SNP. Only six SVs had an $r^2 < 0.5$ with any nearby SNP. For the SV alleles that are significantly associated with traits, all have a SNP in perfect LD. Some of this LD may result from population structure: eight of the nine TEs associated with traits had significant allele frequency differences between populations (Figures S11 and S12).

Despite high LD between SVs and nearby SNPs, many of the associations detected between SVs and traits would not have been captured with a GWAS using all SNPs. Of the four SVs associated with main effects, only one was found in the same peak regions in the SNP GWAS (Figures S13 and S14). This lack of overlap between the SV GWAS associations and the SNP GWAS associations is a result of different significance cutoffs in the two different analyses. The HapMap3 SNP dataset used in the GWAS has 16,435,136 SNPs while there were only 3087 SVs in the SV association mapping analysis, so a SNP needed to have a p-value below 7.94×10^{-6} (averaged across traits) to overcome the FDR cut-off in the SNP GWAS while its linked SV only needed a p-value below 1.86×10^{-4} (averaged across traits) to be detected as significant in the SV GWAS.

4 | Discussion

In this study we leveraged two reference genomes along with a broader set of short-read genomic data to capture SV diversity in a maize diversity panel. The maize genome's highly repetitive nature makes it challenging to rely on short-read alignments alone to characterise SV polymorphism de novo (Hufford et al. 2021). By ascertaining SV presences and absences between two genotypes with long read data, we were able to call SVs across hundreds of maize genotypes using short-read data and identify SVs associated with trait variation.

We found nine SV polymorphisms associated with either average trait value or trait plasticity in a variety of maize phenotypes (Figure 3). Previous studies have identified SVs associated with phenotypic variation that would not be discovered in

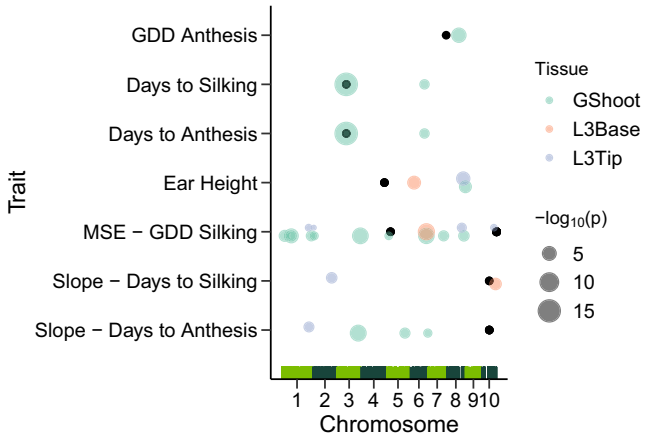

FIGURE 5 | Genomic positions and p-values for genes with expression significantly associated with the genotypes of seven structural variant (SV) markers identified in our genome-wide association analyses—Bars at bottom represent the genomic positions for the 3087 SV markers used in the association panel, with chromosomes in alternating colours. Black points show the position of the SV marker identified in each trait. Coloured points are sized according to the false discovery rate adjusted $-\log_{10}(p)$ with tissue collected from germinating shoot (GShoot) in green, the base of leaf three (L3Base) in orange, and the tip of leaf three (L3Tip) in blue. The SV marker on chromosome three was the most proximal to the identified SV marker, but was still 911 kb away (GDD: growing degree days; MSE: mean squared error).

analyses that use SNPs alone (Yang et al. 2019; Guo et al. 2020; Hartmann 2022; Zhang et al. 2024). Here, while the SV GWAS identified hits that were not present in the SNP GWAS, all SV associations detected were in perfect linkage disequilibrium with SNPs. We did not detect associations that were not captured by the SNP dataset but instead these SVs reached statistical significance because there were many fewer SVs than SNPs. Previous work investigating TE polymorphism in a different maize genetic diversity panel did find that 20% of TEs were not in LD with SNPs but these SNPs tended to be at a low MAF in the population (Qiu et al. 2021). By focusing on common SV polymorphisms we likely have missed many SVs that are low frequency and not in LD with surrounding SNPs—however these low frequency SVs would be unlikely to be associated with trait variation in a GWAS.

Population structure within the Buckler–Goodman panel could also potentially contribute to high LD between TEs associated with traits and nearby SNPs. Although LD tends to decay rapidly in maize, high LD has been observed around multiple QTLs (Flint-Garcia, Thornsberry, and Buckler 2003). For example, previous work in the Buckler–Goodman panel has found long stretches of linkage disequilibrium around a flowering-time associated allele (Hung et al. 2012). For many of the TEs identified as associated with traits, there were significant differences in allele frequencies between maize subpopulations. This pattern is consistent with observations that population structure can contribute a significant amount to phenotypic variation—over 30% for life history traits like days to pollen and days to silking (Flint-Garcia et al. 2005).

Of the SVs included in this study, 91% contained TEs or are themselves of TE origin and the largest category of SVs were

clear examples of TE insertions (21,103 or 23.5%). All but one of the SVs associated with trait variation and with GxE contained TE sequence, yet only the SVs on chromosome 10 for the slope of days to silking and the MSE of growing degree days to silking FW models appeared to be the direct result of TE insertions. The remaining seven associations result from deletions that contain TEs. This result is consistent with previous findings that deletions have been the dominant contributors to SV polymorphism in maize (Munasinghe et al. 2023). We did observe that the SV associated with the MSE of growing degree days to silking on chromosome 10 that appeared to result from a TE insertion was the SV with the most associations with gene expression. This pattern is consistent with hypotheses that TE insertions are particularly likely to affect gene expression (Klein and Anderson 2022), although further work is clearly needed to evaluate how broad this pattern is across a larger sample of SVs.

We found five significant associations between SVs and plasticity, quantified using MSE and slopes from the Finlay–Wilkinson regression models. The finding that different SVs were associated with traits than with trait plasticity is consistent with most previous work. For example, the genetic architecture of trait means and trait plasticity have been shown to differ in maize (Kusmec et al. 2017; Tibbs-Cortes et al. 2024) and *Arabidopsis thaliana* (Fournier-Level et al. 2022) but not sorghum (Wei et al. 2024). We also did not see a clear pattern that SVs are more likely to affect trait variation across environments than trait means, but this may result from having a small number of associations across both categories.

Overall, we have demonstrated an approach for using two reference genomes to identify SVs and then genotype for these variants in a larger panel of individuals with short-read sequencing data. This approach identifies SVs associated with phenotypic variation and with GxE interactions. However, this approach does bias us towards common alleles that were polymorphic within the two reference assemblies. This bias is acceptable for a GWAS, where we will also be biased towards detecting associations with variants at intermediate allele frequency, but would be less appropriate for any analysis that would need to identify SVs with low allele frequencies. As long-read data becomes more affordable and more reference genomes become available for more species, these types of approaches will improve our ability to detect SVs and investigate their potential functional importance.

Author Contributions

Nathan S. Catlin: designed research, performed research, analyzed data, wrote the paper. **Husain I. Agha:** designed research, performed research, analyzed data, wrote the paper. **Adrian E. Platts:** designed research, performed research, analyzed data, reviewed the paper. **Manisha Munasinghe:** designed research, performed research, analyzed data, reviewed the paper. **Candice N. Hirsch:** designed research, project administration, reviewed the paper. **Emily B. Josephs:** designed research, project administration, wrote the paper.

Acknowledgements

We thank Nathan Springer, Jeff Ross-Ibarra, Michelle Stitzer and Yaniv Brandvain, along with members of the Josephs, Hirsch, Ross-Ibarra, Kaepler and Springer labs and two anonymous reviewers for helpful comments and suggestions.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All sequencing data used in this study are publicly available and generated by previous papers. Genomic short-read data were sequenced for the third generation maize haplotype map (HapMap3) project and are available in the NCBI Sequence Read Archive (SRA) under the BioProject ID PRJNA389800 (Bukowski et al. 2018). The B73 and Oh43 reference genomes Zm-B73-REFERENCE-NAM-5.0 and Zm-Oh43-REFERENCE-NAM-1.0, respectively, are available in the MaizeGDB FTP site (https://download.maizegdb.org). Chain files for the genome builds B73 version 3 (APGv3) to B73 version 4 (B73_RefGen_v4) and B73 version 4 to B73 version 5 (Zm-B73-REFERENCE-NAM-5.0) can be found in gramene.org and maizegdb.org, respectively (Tello-Ruiz, Jaiswal, and Ware 2022, Woodhouse et al. 2021). Phenotypes used in association analyses can be found at https://cbsusrv04.tc.cornell.edu/users/panzea/download.aspx?filegroupid=9. Gene expression data were collected from Kremling et al. 2018. VCF files used in LD analyses were downloaded from https://datacommons.cyverse.org/browse/iplant/home/shared/panzea/hapmap3/hmp321/imputed. All code and scripts produced for this manuscript are archived on Github and Zenodo at https://github.com/nscatlin/SVs_282.

Benefit-Sharing Statement

All code and a table of called TE polymorphisms will be made available, as described above.

References

Ågren, J. A., and S. I. Wright. 2011. "Co-Evolution Between Transposable Elements and Their Hosts: A Major Factor in Genome Size Evolution." *Chromosome Research* 19: 777–786.

Alonge, M., X. Wang, M. Benoit, et al. 2020. "Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato." *Cell* 182: 145–161.

Audano, P. A., A. Sulovari, T. A. Graves-Lindsay, et al. 2019. "Characterizing the Major Structural Variant Alleles of the Human Genome." *Cell* 176: 663–675.

Beló, A., M. K. Beatty, D. Hondred, K. A. Fengler, B. Li, and A. Rafalski. 2010. "Allelic Genome Structural Variations in Maize Detected by Array Comparative Genome Hybridization." *Theoretical and Applied Genetics* 120: 355–367.

Benjamini, Y., and Y. Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B: Methodological* 57: 289–300.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30: 2114–2120.

Bukowski, R., X. Guo, Y. Lu, et al. 2018. "Construction of the Third-Generation *Zea mays* Haplotype Map." *GigaScience* 7: 1–12.

Cameron, D. L., L. Di Stefano, and A. T. Papenfuss. 2019. "Comprehensive Evaluation and Characterisation of Short Read General-Purpose Structural Variant Calling Software." *Nature Communications* 10: 3240.

Cao, J., K. Schneeberger, S. Ossowski, et al. 2011. "Whole-Genome Sequencing of Multiple *Arabidopsis thaliana* Populations." *Nature Genetics* 43: 956–963.

Catlin, N. S., and E. B. Josephs. 2022. "The Important Contribution of Transposable Elements to Phenotypic Variation and Evolution." *Current Opinion in Plant Biology* 65: 102140.

Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee. 2015. "Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets." *GigaScience* 4: 7.

Charlesworth, B., and D. Charlesworth. 1983. "The Population Dynamics of Transposable Elements." *Genetics Research* 42: 1–27.

Chia, J. M., C. Song, P. J. Bradbury, et al. 2012. "Maize hapmap2 Identifies Extant Variation From a Genome in Flux." *Nature Genetics* 44: 803–807.

Cosson, P., L. Sofer, Q. Hien Le, et al. 2010. "Rtm3, Which Controls Long-Distance Movement of Potyviruses, Is a Member of a New Plant Gene Family Encoding a Meprin and Traf Homology Domain-Containing Protein." *Plant Physiology* 154: 222–232.

Danecek, P., J. K. Bonfield, J. Liddle, et al. 2021. "Twelve Years of SAMtools and BCFtools." *GigaScience* 10: giab008.

Darracq, A., C. Vitte, S. Nicolas, et al. 2018. "Sequence Analysis of European Maize Inbred Line f2 Provides New Insights Into Molecular and Chromosomal Characteristics of Presence/Absence Variants." *BMC Genomics* 19: 1–20.

de Oliveira, D. S., M. T. Rosa, C. Vieira, and E. L. Loreto. 2021. "Oxidative and Radiation Stress Induces Transposable Element Transcription in *Drosophila melanogaster*." *Journal of Evolutionary Biology* 34: 628–638.

Di Marzo, M., H. Herrera-Ubaldo, E. Caporali, et al. 2020. "Seedstick Controls Arabidopsis Fruit Size by Regulating Cytokinin Levels and Fruitfull." *Cell Reports* 30: 2846–2857.

Domnguez, M., E. Dugas, M. Benchouaia, et al. 2020. "The Impact of Transposable Elements on Tomato Diversity." *Nature Communications* 11: 4058.

Dong, Z., W. Li, E. Unger-Wallace, J. Yang, E. Vollbrecht, and G. Chuck. 2017. "Ideal Crop Plant Architecture Is Mediated by Tassels Replace Upper Ears1, a Btb/Poz Ankyrin Repeat Gene Directly Targeted by Teosinte Branched1." *National Academy of Sciences of the United States of America* 114: E8656–E8664.

Dreni, L., S. Jacchia, F. Fornara, et al. 2007. "The d-Lineage Mads-Box Gene osmads13 Controls Ovule Identity in Rice." *Plant Journal* 52: 690–699.

Ebert, P., P. A. Audano, Q. Zhu, et al. 2021. "Haplotype-Resolved Diverse Human Genomes and Integrated Analysis of Structural Variation." *Science* 372: eabf7117.

Finlay, K., and G. Wilkinson. 1963. "The Analysis of Adaptation in a Plant-Breeding Programme." *Australian Journal of Agricultural Research* 14: 742–754.

Flint-Garcia, S. A., J. M. Thornsberry, and E. S. Buckler. 2003. "Structure of Linkage Disequilibrium in Plants." *Annual Review of Plant Biology* 54: 357–374.

Flint-Garcia, S. A., A. C. Thuillet, J. Yu, et al. 2005. "Maize Association Population: A High-Resolution Platform for Quantitative Trait Locus Dissection." *Plant Journal* 44: 1054–1064.

Fournier-Level, A., M. A. Taylor, J. F. Paril, et al. 2022. "Adaptive Significance of Flowering Time Variation Across Natural Seasonal Environments in *Arabidopsis thaliana*." *New Phytologist* 234: 719–734.

Franco, J. A. V., J. L. Gage, P. J. Bradbury, et al. 2020. "A Maize Practical Haplotype Graph Leverages Diverse Nam Assemblies. *bioRxiv*." https://doi.org/10.1101/2020.08.31.268425.

Fray, R. G., and D. Grierson. 1993. "Identification and Genetic Analysis of Normal and Mutant Phytoene Synthase Genes of Tomato by Sequencing, Complementation and Co-Suppression." *Plant Molecular Biology* 22: 589–602.

Fueyo, R., J. Judd, C. Feschotte, and J. Wysocka. 2022. "Roles of Transposable Elements in the Regulation of Mammalian Transcription." *Nature Reviews Molecular Cell Biology* 23: 481–497.

Gage, J. L., B. Vaillancourt, J. P. Hamilton, et al. 2019. "Multiple Maize Reference Genomes Impact the Identification of Variants by Genome-Wide Association Study in a Diverse Inbred Panel." *Plant Genome* 12: 180069.

Gui, S., W. Wei, C. Jiang, et al. 2022. "A Pan-Zea Genome Map for Enhancing Maize Improvement." *Genome Biology* 23: 178.

Guo, J., K. Cao, C. Deng, et al. 2020. "An Integrated Peach Genome Structural Variation Map Uncovers Genes Associated With Fruit Traits." *Genome Biology* 21: 1–19.

Hämälä, T., E. K. Wafula, M. J. Guiltinan, P. E. Ralph, C. W. Depamphilis, and P. Tiffin. 2021. "Genomic Structural Variants Constrain and Facilitate Adaptation in Natural Populations of *theobroma cacao*, the Chocolate Tree." *National Academy of Sciences of the United States of America* 118: e2102914118.

Hartmann, F. E. 2022. "Using Structural Variants to Understand the Ecological and Evolutionary Dynamics of Fungal Plant Pathogens." *New Phytologist* 234: 43–49.

Hirsch, C. D., and N. M. Springer. 2017. "Transposable Element Influences on Gene Expression in Plants." *Biochimica et Biophysica Acta* 1860: 157–165.

Hof, A. E., P. Campagne, D. J. Rigden, et al. 2016. "The Industrial Melanism Mutation in British Peppered Moths Is a Transposable Element." *Nature* 534: 102–105.

Huddleston, J., M. J. Chaisson, K. M. Steinberg, et al. 2017. "Discovery and Genotyping of Structural Variation From Long-Read Haploid Genome Sequence Data." *Genome Research* 27: 677–685.

Hufford, M. B., A. S. Seetharam, M. R. Woodhouse, et al. 2021. "De Novo Assembly, Annotation, and Comparative Analysis of 26 Diverse Maize Genomes." *Science* 373, no. 6555: 655–662. https://doi.org/10.1126/science.abg5289.

Hung, H. Y., L. M. Shannon, F. Tian, et al. 2012. "Zmcct and the Genetic Basis of Day-Length Adaptation Underlying the Postdomestication Spread of Maize." *National Academy of Sciences of the United States of America* 109: E1913–E1921.

Juranić, M., S. Ko, N. G. Krohn, D. Leljak-Levanić, S. Sprunck, and T. Dresselhaus. 2012. "Germline-Specific Math-Btb Substrate Adaptor mab1 Regulates Spindle Length and Nuclei Identity in Maize." *Plant Cell* 24: 4974–4991.

Kang, M., H. Wu, H. Liu, et al. 2023. "The Pan-Genome and Local Adaptation of *Arabidopsis thaliana*." *Nature Communications* 14: 6259.

Klein, S. P., and S. N. Anderson. 2022. "The Evolution and Function of Transposons in Epigenetic Regulation in Response to the Environment." *Current Opinion in Plant Biology* 69: 102277.

Kobayashi, S., N. Goto-Yamamoto, and H. Hirochika. 2004. "Retrotransposon-Induced Mutations in Grape Skin Color." *Science* 304: 982.

Kremling, K. A. G., S. Y. Chen, M. H. Su, et al. 2018. "Dysregulation of Expression Correlates With Rare-Allele Burden and Fitness Loss in Maize." *Nature* 555: 520–523.

Kusmec, A., S. Srinivasan, D. Nettleton, and P. S. Schnable. 2017. "Distinct Genetic Architectures for Phenotype Means and Plasticities in *Zea mays*." *Nature Plants* 3: 715–723.

Li, H., W. Liang, Y. Hu, et al. 2011. "Rice mads6 Interacts With the Floral Homeotic Genes superwoman1, mads3, mads58, mads13, and Drooping Leaf in Specifying Floral Organ Identities and Meristem Fate." *Plant Cell* 23: 2536–2552.

Liang, Z., S. N. Anderson, J. M. Noshay, P. A. Crisp, T. A. Enders, and N. M. Springer. 2021. "Genetic and Epigenetic Variation in Transposable Element Expression Responses to Abiotic Stress in Maize." *Plant Physiology* 186: 420–433.

Lisch, D. 2013. "How Important Are Transposons for Plant Evolution." *Nature Reviews Genetics* 14: 49–61.

Liu, Y., H. Du, P. Li, et al. 2020. "Pan-Genome of Wild and Cultivated Soybeans." *Cell* 182: 162–176.

Lopez-Dee, Z. P., P. Wittich, M. Enrico Pe, et al. 1999. "Osmads13, a Novel Rice Mads-Box Gene Expressed During Ovule Development." *Developmental Genetics* 25: 237–244.

Love, M., S. Anders, and W. Huber. 2014. "Differential Analysis of Count Data–the deseq2 Package." *Genome Biology* 15: 10–1186.

Lucas Lledó, J. I., and M. Cáceres. 2013. "On the Power and the Systematic Biases of the Detection of Chromosomal Inversions by Paired-End Genome Sequencing." *PLoS One* 8: e61292.

Makarevitch, I., A. J. Waters, P. T. West, et al. 2015. "Transposable Elements Contribute to Activation of Maize Genes in Response to Abiotic Stress." *PLoS Genetics* 11: e1004915.

Mei, W., M. G. Stetter, D. J. Gates, M. C. Stitzer, and J. Ross-Ibarra. 2018. "Adaptation in Plant Genomes." *American Journal of Botany* 105: 16–19.

Milyaeva, P. A., I. V. Kukushkina, A. I. Kim, and L. N. Nefedova. 2023. "Stress Induced Activation of Ltr Retrotransposons in the *Drosophila melanogaster* Genome." *Life (Bethesda)* 13: 2272.

Mitchell-Olds, T., J. H. Willis, and D. B. Goldstein. 2007. "Which Evolutionary Processes Influence Natural Genetic Variation for Phenotypic Traits?" *Nature Reviews Genetics* 8: 845–856.

Mizzotti, C., I. Ezquer, D. Paolo, et al. 2014. "Seedstick Is a Master Regulator of Development and Metabolism in the Arabidopsis Seed Coat." *PLoS Genetics* 10: e1004856.

Munasinghe, M., A. Read, M. C. Stitzer, et al. 2023. "Combined Analysis of Transposable Elements and Structural Variation in Maize Genomes Reveals Genome Contraction Outpaces Expansion." *PLoS Genetics* 19, no. 12: e1011086. https://doi.org/10.1371/journal.pgen.1011086.

Oelmüller, R., T. Peškan-Berghöfer, B. Shahollari, A. Trebicka, I. Sherameti, and A. Varma. 2005. "Math Domain Proteins Represent a Novel Protein Family in Arabidopsis Thaliana, and at Least One Member Is Modified in Roots During the Course of a Plant–Microbe Interaction." *Physiologia Plantarum* 124: 152–166.

Park, M., S. G. Hong, H. Park, B. Lee, and H. Lee. 2018. "Identification of Reference Genes for Rt-Qpcr in the Antarctic Moss *Sanionia uncinata* Under Abiotic Stress Conditions." *PLoS One* 13: e0199356.

Peiffer, J. A., M. C. Romay, M. A. Gore, et al. 2014. "The Genetic Architecture of Maize Height." *Genetics* 196: 1337–1356.

Pic. 2019. "Picard Toolkit." https://broadinstitute.github.io/picard/.

Qin, P., H. Lu, H. Du, et al. 2021. "Pan-Genome Analysis of 33 Genetically Diverse Rice Accessions Reveals Hidden Genomic Variations." *Cell* 184: 3542–3558.

Qiu, Y., C. H. O'Connor, R. Della Coletta, et al. 2021. "Whole-Genome Variation of Transposable Element Insertions in a Maize Diversity Panel." *G3: Genes, Genomes, Genetics* 11: jkab238.

Quinlan, A. R., and I. M. Hall. 2010. "Bedtools: A Flexible Suite of Utilities for Comparing Genomic Features." *Bioinformatics* 26: 841–842.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. "Edger: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data." *Bioinformatics* 26: 139–140.

Schatz, M. 2018. "Michael Schatz: 100 Genomes in 100 Days: The Structural Variant Landscape in Tomato Genomes. Nanopore Community Meeting 2018".

Servin, B., and M. Stephens. 2007. "Imputation-Based Analysis of Association Studies: Candidate Regions and Quantitative Traits." *PLoS Genetics* 3: e114.

Shimazaki, M., K. Fujita, H. Kobayashi, and S. Suzuki. 2011. "Pink-Colored Grape Berry Is the Result of Short Insertion in Intron of Color Regulatory Gene." *PLoS One* 6: e21308.

Sieber, A. N., C. F. H. Longin, W. L. Leiser, and T. Würschum. 2016. "Copy Number Variation of Cbf-a14 at the Fr-a2 Locus Determines Frost Tolerance in Winter Durum Wheat." *Theoretical and Applied Genetics* 129: 1087–1097.

Sirén, J., N. Välimäki, and V. Mäkinen. 2014. "HISAT2-Fast and Sensitive Alignment Against General Human Population." *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 11: 375–388.

Song, B., S. Marco-Sola, M. Moreto, L. Johnson, E. S. Buckler, and M. C. Stitzer. 2022. "AnchorWave: Sensitive Alignment of Genomes With High Sequence Diversity, Extensive Structural Polymorphism, and Whole-Genome Duplication." *Proceedings of the National Academy of Sciences of the United States of America* 119: e2113075119.

Springer, N. M., K. Ying, Y. Fu, et al. 2009. "Maize Inbreds Exhibit High Levels of Copy Number Variation (Cnv) and Presence/Absence Variation (Pav) in Genome Content." *PLoS Genetics* 5: e1000734.

Studer, A., Q. Zhao, J. Ross-Ibarra, and J. Doebley. 2011. "Identification of a Functional Transposon Insertion in the Maize Domestication Gene tb1." *Nature Genetics* 43: 1160–1163.

Sun, L., Y. Jing, X. Liu, et al. 2020. "Heat Stress-Induced Transposon Activation Correlates With 3d Chromatin Organization Rearrangement in Arabidopsis." *Nature Communications* 11: 1886.

Tello-Ruiz, M. K., P. Jaiswal, and D. Ware. 2022. "Gramene: A Resource for Comparative Analysis of Plants Genomes and Pathways." *Methods in Molecular Biology* 2443: 101–131.

Tibbs-Cortes, L. E., T. Guo, C. M. Andorf, X. Li, and J. Yu. 2024. "Comprehensive Identification of Genomic and Environmental Determinants of Phenotypic Plasticity in Maize." *Genome Research* 34: 1253–1263.

Van der Knaap, E., A. Sanyal, S. Jackson, and S. Tanksley. 2004. "High-Resolution Fine Mapping and Fluorescence In Situ Hybridization Analysis of Sun, a Locus Controlling Tomato Fruit Shape, Reveals a Region of the Tomato Genome Prone to Dna Rearrangements." *Genetics* 168: 2127–2140.

Walley, J. W., R. C. Sartor, Z. Shen, et al. 2016. "Integration of Omic Networks in a Developmental Atlas of Maize." *Science* 353: 814–818.

Wang, X., L. Gao, C. Jiao, et al. 2020. "Genome of *Solanum pimpinellifolium* Provides Insights Into Structural Variants During Tomato Breeding." *Nature Communications* 11: 5817.

Wang, Y., Y. Liu, S. Qu, et al. 2022. "Nitrogen Starvation Induces Genome-Wide Activation of Transposable Elements in Arabidopsis." *Journal of Integrative Plant Biology* 64: 2374–2384.

Wei, J., T. Guo, Q. Mu, et al. 2024. "Genetic and Environmental Patterns Underlying Phenotypic Plasticity in Flowering Time and Plant Height in Sorghum." *Plant, Cell & Environment*. https://doi.org/10.1111/pce.15213.

Woodhouse, M. R., E. K. Cannon, J. L. Portwood, et al. 2021. "A Pan-Genomic Approach to Genome Databases Using Maize as a Model System." *BMC Plant Biology* 21: 1–10.

Xu, X., X. Liu, S. Ge, et al. 2012. "Resequencing 50 Accessions of Cultivated and Wild Rice Yields Markers for Identifying Agronomically Important Genes." *Nature Biotechnology* 30: 105–111.

Yang, N., J. Liu, Q. Gao, et al. 2019. "Genome Assembly of a Tropical Maize Inbred Line Provides Insights Into Structural Variation and Crop Improvement." *Nature Genetics* 51: 1052–1059.

Zhang, Z., J. P. G. Viana, B. Zhang, et al. 2024. "Major Impacts of Widespread Structural Variation on Sorghum." *Genome Research* 34: 286–299.

Zhou, X., and M. Stephens. 2012. "Genome-Wide Efficient Mixed-Model Analysis for Association Studies." *Nature Genetics* 44: 821–824.

Zhou, Y., A. Minio, M. Massonnet, et al. 2019. "The Population Genetics of Structural Variants in Grapevine Domestication." *Nature Plants* 5: 965–979.

Zhou, Y., Z. Zhang, Z. Bao, et al. 2022. "Graph Pangenome Captures Missing Heritability and Empowers Tomato Breeding." *Nature (London, England)* 606: 527–534.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.