Domestication Reshaped the Genetic Basis of Inbreeding Depression in a Maize Landrace Compared to its Wild Relative, Teosinte

L.F. Samayoa[*], B.A. Olukolu[†], C.J. Yang[‡], Q. Chen[‡], Markus G. Stetter[§], Alessandra M. York[‡], Jose de Jesus Sanchez-Gonzalez[**], Jeffrey C. Glaubitz[††], Peter J. Bradbury[‡‡], Maria Cinta Romay[††], Qi Sun[††], Jinliang Yang[§§], Jeffrey Ross-Ibarra[***], Edward S. Buckler[‡‡], J.F. Doebley[‡], and J.B. Holland[*,†††]

[*]Department of Crop and Soil Sciences, North Carolina State University, Raleigh, NC 27695 USA

[†]Department of Entomology and Plant Pathology, University of Tennessee, Knoxville, Tennessee 37996 USA

[‡] Laboratory of Genetics, University of Wisconsin – Madison, Madison, WI 53706 USA

[§] Institute for Plant Sciences and Center of Excellence on Plant Sciences, University of Cologne, Cologne 50674 Germany

[**] Centro Universitario de Ciencias Biológicas y Agropecuarias, Universidad de Guadalajara, Zapopan, Jalisco CP45110, Mexico

[††] Institute of Biotechnology, Cornell University, Ithaca, NY 14853 USA

[‡‡] US Department of Agriculture–Agricultural Research Service, Cornell University, Ithaca, NY 14853 USA

[§§] Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE 68583 USA

[***] Department of Evolution and Ecology, Center for Population Biology, and Genome Center, University of California, Davis, CA 95616 USA

[†††] United States Department of Agriculture – Agriculture Research Service, Raleigh, NC 27695 USA

Abstract

Inbreeding depression is the reduction in fitness and vigor resulting from mating of close relatives observed in many plant and animal species. The extent to which the genetic load of mutations contributing to inbreeding depression is due to rare large-effect variation versus potentially more common variants with very small individual effects is unknown and may be affected by population history. We compared the effects of outcrossing and self-fertilization on 18 traits in a landrace population of maize, which underwent a population bottleneck during domestication, and a neighboring population of its wild relative teosinte. Inbreeding depression was greater in maize than teosinte for 15 of 18 traits, congruent with the greater segregating genetic load predicted from sequence data in the maize population. For many traits - and more commonly in maize - genetic variation among self-fertilized families was less than expected based on additive and dominance variance estimated in outcrossed families, suggesting that a negative covariance between additive and homozygous dominance effects limits the variation available to selection under partial inbreeding. We identified quantitative trait loci (QTL) representing large-effect rare variants carried by only a single parent, which were more important in teosinte than maize. Teosinte also carried more putative juvenile-acting lethal variants identified by segregation distortion. These results suggest a mixture of mostly polygenic, small-effect recessive variation underlying inbreeding depression, with an additional contribution from rare larger-effect variants that was more important in teosinte but depleted in maize following to the domestication bottleneck. Purging associated with the maize domestication bottleneck may have selected against large effect variants, but polygenic load is harder to purge and segregating mutational burden increased in maize compared to teosinte.

Introduction

Darwin (1876) demonstrated experimentally the detrimental effect of self-fertilization in numerous plant species, including an average reduction of 17% in the height of adult maize (*Zea mays* ssp. *mays*) plants derived from self-fertilization compared to outcrossing. The reduction in vigor and fitness of plants due to self-fertilization is an extreme example of inbreeding depression observed across many species, including humans, resulting from matings of close relatives. The genetic basis for inbreeding depression is largely due to at least partially recessive deleterious variants whose effects are expressed when homozygous (Charlesworth and Charlesworth 1999). Inbreeding depression depends on the level and direction of genetic dominance as well as the frequency of deleterious variants, being maximum when these variants are at equal frequency to favorable variants (Falconer and Mackay 1996). Partial dominance of favorable alleles appears to be the primary gene action affecting fitness in wild species and yield in maize (Charlesworth and Willis 2009; Yang *et al.* 2017a), although overdominance at a subset of loci contributing a small amount of variation cannot be ruled out. Repulsion-phase linkage and low recombination between recessive deleterious mutations can generate pseudo-overdominance or associative overdominance at the level of haplotypes, contributing to the retention of deleterious mutations (Becher *et al.* 2020; Gilbert *et al.* 2020).

A wide range of susceptibility to inbreeding depression is observed among plant species, with some species having evolved mating systems with high levels of self-fertilization accompanied by little inbreeding depression (Charlesworth and Charlesworth 1987, 1999; Husband and Schemske 1996). The ability of some species to tolerate high levels of self-fertilization is evidence for purging of recessive deleterious alleles during their evolution (Charlesworth and Charlesworth 1999). Variation for inbreeding depression also exists within species, from the sub-population to the level of individual parents (Schultz and Willis 1995; Fowler and Whitlock 1999; Waller *et al.* 2008; Escobar *et al.* 2008). Intraspecific variation for inbreeding depression impacts the evolution of mating systems (Kelly 2005a) and the response to selection in populations with mixed outcrossing and selfing (Edwards 2008).

Recent research has begun to identify the genomic basis underpinning inbreeding depression, revealing some large-effect recessive variants segregating in natural and breeding populations (Howard *et al.* 2017; Zhang *et al.* 2019), substantial loss of transposable elements in some lineages (Roessler *et al.* 2019), and genome-wide alterations in gene expression due to inbreeding (Paige 2010; Swanson-Wagner *et al.* 2012). Some important questions about inbreeding depression remain unresolved, however. How much inbreeding depression is due to lethal or large-effect recessive mutations vs. polygenic small effect variants distributed widely throughout the genome (Charlesworth and Willis 2009)? Although directional selection is expected to be sufficiently strong to eliminate much of the recessive genetic load from outcrossing populations, why is inbreeding depression for fitness-related traits so severe in many outcrossing species (Charlesworth and Charlesworth 1999; Charlesworth and Willis 2009; Troth *et al.* 2018)? How much inbreeding depression is expressed as lethality early in development vs. reduced adult vigor and fitness (Husband and Schemske 1996)? Do the additional components of covariances among relatives introduced by inbreeding limit the effectiveness of selection in mixed outcrossing and selfing species like maize (Cockerham and Weir 1984; Shaw *et al.* 1998; Edwards and Lamkey 2002; Moorad and Wade 2005)?

The responses of genetic load and inbreeding depression to population bottlenecks are complex (Simons *et al.* 2014). Reduced population size and genetic drift can increase the frequency of some rare variants to fixation, resulting in an increase in the fixed genetic load, which reduces the mean fitness of populations independently of their level of inbreeding. In contrast, inbreeding depression due to segregating genetic load is expected to be reduced following bottlenecks due to the enhanced efficiency of purging selection in small populations (Lande and Schemske 1985; Kirkpatrick and Jarne 2000). The different demographic histories of maize landraces and teosinte are expected to result in differences in their genetic load of deleterious alleles. Wang et al. (2017) and Lozano et al. (2021) reported that maize has higher predicted deleterious genomic burden than teosinte, estimated based on evolutionary conservation (Davydov *et al.* 2010) or the predicted effects of amino acid substitutions (Vaser *et al.* 2016) of SNPs. Importantly, Wang et al. (2017) demonstrated that maize has more fixed putative deleterious sites than teosinte, but such fixed deleterious variants reduce the fitness of both outbred and inbred progenies equally, thus they do not contribute to inbreeding depression. In contrast, maize has fewer segregating predicted deleterious sites than teosinte (Wang *et al.* 2017), leading to the expectation that maize should exhibit less inbreeding depression than teosinte (Waller 2021). Evidence for the functional importance of this predicted segregating load in maize is that hybrid vigor in crosses between elite maize inbred lines can be largely attributed to complementation of the predicted SNP burdens of different inbred parents (Yang *et al.* 2017a).

Large effect variants that reduce fitness are selected against, but may be maintained in populations at low frequency if they are recessive, masked in heterozygotes by more favorable dominant alleles (Charlesworth and Willis 2009). This hinders the study of genes underlying inbreeding depression, as large-effect deleterious variants occur at low frequency, and variants at higher frequency have only small effects, limiting power of detection and estimation of their effects by genome-wide association study (Gibson 2012). An alternative approach to the detection of specific variants is to test the effects of groups of variants classified by population genetic or evolutionary signatures associated with deleterious effects (Karczewski *et al.* 2020). For example, Brown and Kelly (2020) demonstrated that the proportion of rare alleles (the "rare allele load") carried by inbred *Mimulus* lines was correlated to reduced fitness. In maize inbreds, rare alleles tend to have dysregulated gene expression, which is correlated with lower seed yield (Kremling *et al.* 2018). Variants in evolutionarily conserved sites tend to be rare in maize inbreds and complementation of such variants by common alleles in hybrids is associated with higher yield (Yang *et al.* 2017a).

Maize was domesticated from the wild grass teosinte (*Zea mays* ssp. *parviglumis*) in southwestern Mexico about 9,000 years ago (Matsuoka *et al.* 2002; Piperno *et al.* 2009), and the extant ranges of the two subspecies overlap in Mexico and Central America (Sánchez González *et al.* 2018). Following domestication, maize experienced a population bottleneck (Eyre-Walker *et al.* 1998; Beissinger *et al.* 2016), human selection for traits associated with easier harvest and consumption (Hernández Xolocotzi 1985; Yang *et al.* 2019), and introgression from a weedy relative, (*Zea mays* ssp. *mexicana*)(Hufford *et al.* 2013; Yang *et al.* 2017b). The shared evolutionary lineage and environmental adaptation of maize and teosinte populations that still inhabit a common range permit a comparison of

the evolution of inbreeding depression in response to their distinct post-domestication selection pressures and demographies.

Maize and teosinte are both predominantly outcrossing species that also naturally self-fertilize at a low rate (Hufford *et al.* 2011, 2012b). Modern maize hybrids are created by first selecting highly homozygous inbred lines, followed by crossing unrelated to restore heterozygosity across much of the genome, resulting in hybrid vigor (Holland 2009). Selection during inbreeding and among resulting inbred lines provides ample opportunity for purging large effect deleterious variants as well as polygenic genetic load, since commercial inbreds are selected for relatively high levels of pollen or seed production (Goodman *et al.* 2014). The rate of genetic gain for yield in maize hybrids was greater than in maize inbreds, resulting in an increase in the amount of inbreeding depression over generations of yield selection in modern maize (Meghji *et al.* 1984; Lamkey and Smith 1987). In contrast to commercial hybrids, maize landraces are maintained as open-pollinated populations selected for local adaptation as well as for ear and kernel type, although migration among populations arising from seed exchange among farmers is also common (Bellon and Brush 1994; Pressoir and Berthaud 2004; Bellon *et al.* 2018; Samayoa *et al.* 2018). Similarly, teosinte populations are predominantly outcrossing; however, reductions in teosinte habitat area may result in inbreeding due to small effective population sizes (Pyhäjärvi *et al.* 2013; Sánchez González *et al.* 2018).

In this study we investigated the genetic architecture of inbreeding depression in maize and teosinte by self- and cross-mating samples of 40 and 49 outbred plants sampled from a maize landrace and a teosinte population, respectively, collected from nearby locations. The resulting large family sizes provide good power for estimating phenotypic effects of variants carried by those parents, even if the variants are rare in the larger populations from which the parents were sampled. We documented the magnitude of inbreeding depression and changes in genetic variance under inbreeding for a range of traits in both teosinte and maize. We compared the two populations for predicted segregating genetic load based on evolutionary conservation of SNP variants. We mapped the effects of rare alleles private to individual founders, measured the relative importance of polygenes versus detectable QTL using a novel mapping strategy (Figure S1), identified putative juvenile-stage lethal alleles based on linked segregation distortion, and estimated the phenotypic effects of parental rare allele burden in the maize and teosinte populations to understand how domestication and the different histories of these populations influenced inbreeding depression and its genetic architecture.

We also compared the effects of self-fertilization on the trait genotypic variance among selfed families in maize and teosinte. Although genotypic variance in outbred populations is composed of additive, dominant, and epistatic genetic variances (Falconer and Mackay 1996), additional genetic covariances impact the genotypic variance under inbreeding (Cockerham and Weir 1984; Walsh and Lynch 2018). These additional components can impact the response to natural or artificial selection in populations with mixed outcrossing and selfing, such as maize and teosinte. Our mating design permitted estimation of additive and dominance variances in outbred individuals. Using these estimates, quantitative genetic theory permits prediction of the total genotypic variance among selfed family means under the assumption that the additive and dominance components are the only meaningful contributors to the inbred family genotypic variance. Deviations between the observed and predicted

variation among selfed family means indicate the extent to which selection among outbred individuals contributes to reduction or increase in inbreeding depression (Edwards 2008).

RESULTS AND DISCUSSION

Maize exhibits more inbreeding depression than teosinte

We evaluated 4,455 teosinte plants grouped into 49 selfed and 377 outcross full-sib families, and 4,398 maize landrace plants grouped into 34 selfed and 89 outcross full-sib families for 18 traits related to domestication (Tables S1 and S2; Yang *et al.* 2019). The 18 measured traits were grouped into vegetative/flowering time, environmental response, and reproductive categories (Table 1; Yang *et al.* 2019). The selfed and outcrossed progeny means were estimated while accounting for parental relationships and non-genetic effects using a diallel model. The coefficient of inbreeding depression measured as the absolute proportional change in population mean from a single generation of selfing was higher on average for environmental response traits within both subspecies, 21% (in teosinte) and 36% (in maize; (Figure 1; Figure S2; Table S3). The absolute value of population mean change due to selfing averaged about 10% in both populations for vegetative traits and 11% in teosinte and 32% in maize for reproductive traits. However, the high level of inbreeding depression for environmental response traits was strongly affected by tiller number (TILN), which had a mean value near zero in outbred maize (and very little variation), such that the small numeric change in TILN due to selfing corresponds to a large percentage change. Removing TILN from the mean inbreeding depression effect calculations reduces the mean selfing effect on environmental response traits to 19% in teosinte and to 20% in maize, which is less than the mean inbreeding depression for reproductive traits (Table S3). Maize had greater inbreeding depression than teosinte for 15 of 18 traits, for each trait category mean, and had about twice as much inbreeding depression for reproductive traits.

The higher inbreeding depression in maize for reproductive traits is initially surprising given the lower number of segregating deleterious variants in maize predicted on the basis of evolutionary conservation by Wang *et al.* (2017). To investigate patterns of deleterious variation predicted from sequence conservation alone, we used Genomic Evolutionary Rate Profiling (GERP) scores calculated by Kistler *et al*. (2018) to identify putatively deleterious SNPs in our populations. Congruent with Wang *et al*. (2017) and with the higher diversity within teosinte, we identified a total of 293,720 vs. 362,145 variants with GERP scores segregating within maize and teosinte, respectively. The number of segregating deleterious sites is not the sole determinant of inbreeding depression, however, because mutations vary for their predicted effects on fitness and for allele frequency. While the total number of deleterious SNPs in the population is higher in teosinte, the number of sites at which a given parent carries a predicted deleterious variant is higher in maize (Figure 2A; Figure S3; Table S4). Taking into account the predicted magnitude and zygosity of each variant, the mean burden of each of these sites is slightly lower in maize (Figure 2B). Combined, the total predicted mutational burden per parent, which depends on both the number and effect of deleterious alleles carried by each parent, is higher in maize than in teosinte (Figure 2C; Figure S2, Table S4), congruent with the generally greater inbreeding depression observed in maize. Taken together, these observations suggest that individual maize plants tend to carry a larger number of more weakly deleterious variants. Consistent with this, the average frequency of derived alleles is higher in maize compared to teosinte (Figure 2D).

The 15% decrease in height and the increase in days to flowering resulting from selfing in the maize landrace is very similar to that reported by Hallauer and Sears (1973) and Cornelius and Dudley

(1974) using open-pollinated populations in the North Central corn belt region of the USA. The more than 50% decrease in total grain weight per plant (TGWP) due to selfing observed here is greater than the about 40% decrease reported in the earlier studies, however. The population studied by Hallauer and Sears (1973) was founded by inbred lines, thus excluding the possibility of segregating recessive lethal alleles. It is likely that landrace populations of maize are segregating for more deleterious variants than are breeding populations that have been selected for productivity using randomized and replicated family-based breeding evaluations. Landrace maize is often selected for very specific ear and kernel forms and culinary uses (Hernández Xolocotzi 1985; Bellon and Brush 1994), and this may also reduce selection on traits related to productivity and yield in landraces.

Linear and non-linear responses to inbreeding

As an alternative to the comparison of selfed and outcrossed progeny means from the diallel analysis, we also regressed trait values on the genomic inbreeding coefficient (β_F) estimated from markers while accounting for relatedness among the individuals with genomic relationship matrices (Yang *et al.* 2019). This estimate accounts for the fine-scale variation in realized inbreeding values among progenies (Figure S4). The linear regression coefficients were highly correlated with (r = 0.99 in teosinte and 0.94 in maize) and almost exactly twice the coefficient of inbreeding (Figure 1; Table S3). This accords with expectation, because one generation of selfing is expected to cause inbreeding to 50% of complete homozygosity. The proportion of genetic variance due to dominance estimated from realized relationship matrices by Yang et al. (2019) was also positively correlated with β_F for each trait (r = 0.68 in teosinte, r = 0.63 in maize; Figure 1, Tables S5 and S6).

We also tested for quadratic regression of trait values on inbreeding coefficients in this framework as a test for epistasis underlying inbreeding depression (Falconer and Mackay 1996; Lynch and Walsh 1998). The quadratic regression coefficient was significant ($p < 0.05$) for plant height (PLHT), leaf width (LFWD), lateral branch length (LBLN), and lateral branch internode length (LBIL) in teosinte but not for any trait in maize. We infer that epistasis was not an important component of the genetic architecture of inbreeding depression for most traits in teosinte and all traits in maize. For the four traits where epistasis appears to be an important component of inbreeding depression, the regression slopes for the quadratic terms were all negative (in the same direction as the linear term), indicating that inbreeding depression accelerated with increasing homozygosity. This is evidence of synergistic effects among deleterious mutations, which may arise due to genetic complementation (Crow and Kimura 1970; Kelly 2005b). For example, a deleterious mutation at one locus may be masked by functional alleles at complementary loci, so that genomic duplication buffers against the early generations of inbreeding, but this buffering can break down as the probability of simultaneous homozygosity at complementarily interacting loci increases at higher levels of inbreeding. The loss of epistasis as a component of inbreeding depression for PLHT, LFWD, LBLN, and LBIL in maize could be the result of the domestication bottleneck. Yang et al. (2019) previously demonstrated a widespread reduction in additive genetic variance when comparing the maize and teosinte populations studied here. The

reduced effective population size associated with the domestication bottleneck is expected to decrease epistatic variance even more strongly than additive variance (Walsh and Lynch 2018).

Reciprocal genetic effects and changes in residual variances associated with inbreeding

The diallel model employed in this study allows estimation of quantitative genetic parameters beyond the coefficient of inbreeding, including additive and dominance genetic variances, reciprocal genetic variances, separate estimates of the micro-environmental residual variances for the selfed and outcrossed progenies, and correlations between parental breeding values in the selfed and outcrossed progenies. Narrow-sense heritability estimates based on additive genetic variances were highly correlated (r = 0.85 in teosinte and 0.83 maize) between the previous analysis of realized relationships and the current diallel model.

Separate estimation of the residual non-genetic variation associated with selfed vs. outcrossed progenies permits a test of the hypothesis that outbred and inbred plants differ for environmental variance (Falconer and Mackay 1996; Whitlock and Fowler 1999). Our results demonstrate that environmental variance increased with inbreeding on average only for maize vegetative and reproductive traits (mean increases of 40% and 20%, respectively), whereas residual environmental variance decreased on average with inbreeding for maize environmental response traits (mean decrease of 39%) and all groups of teosinte traits (decreased from 3% to 29%). The cases of decreased environmental variance may occur due to reduced scale associated with inbreeding, whereas the increased environmental variance observed for maize vegetative and reproductive traits suggests that partially inbred maize plants have reduced capacity for homeostasis.

Reciprocal genetic effects can include cytoplasmic, maternal, and epigenetic parent-of-origin effects (Gonzalo *et al.* 2007), which have been proposed to impact the trajectory of evolution (Räsänen and Kruuk 2007; Lawson *et al.* 2013), including in plants (Galloway *et al.* 2009). Based on model fit criteria, we observed no evidence for reciprocal genetic variance for any trait in either population. This suggests that the potential for maternal and parent-of-origin effects to shape the response to selection within these populations is limited.

A feature of our design is the ability to estimate the breeding values for each parent in both their outcross and inbred progenies (Table S7) and to include the covariance between these breeding values in the model. The covariance of parental breeding values measured in outcrossed versus selfed progenies involves additive variance and a covariance between additive effects and their homozygous dominance deviations, and the variance among selfed progenies includes both of these terms in addition to some of the dominance variance and other higher order covariances (Cockerham and Weir 1984; Walsh and Lynch 2018). Therefore, the correlation of a parent's breeding value between outcrossed and selfed progenies is expected to equal 1 if all of the genetic variance is additive and less than 1 when dominance and other higher order covariances are important. The correlation between parental breeding values across outbred and inbred progenies was $r > 0.90$ for all traits except EL in maize and LBIL in teosinte (Figure 3). These results reflect the fact that even though dominance effects

are important for many of the traits measured (demonstrated by the significant inbreeding depression), most of the genetic variance is additive, which is true across a wide range of genetic architectures (Hill *et al.* 2008).

Parent-specific inbreeding depression is negatively correlated with parental outbred breeding value

We were also able to measure the inbreeding depression for each parent separately by comparing the mean values of their inbred and outbred offspring (correcting for the effects of the other parents on outbred progenies; Figure 3; Table S7). We find very strong negative correlations between the parental breeding value measured in outbred offspring and the family-specific inbreeding depression; except for TILN in maize all correlations were negative, with most less than -0.90 (Table S3). Thus, although the breeding value of a parent was generally highly correlated across progeny types, the amount of family-specific inbreeding depression is reduced for families with higher outbred mean values (Figure 3). These results suggest that most of the genetic load for most traits is due to partially recessive alleles, such that some of the genetic load is expressed in outbred progenies, but the effect of genetic load is exaggerated by inbreeding and increasing the level of homozygosity at loci with unfavorable variants.

The very high correlations between outbred and inbred progeny breeding values measured here contrast with low to moderate correlations estimated for grain yield, plant height, and grain weight from breeding populations by Cornelius and Dudley (1974) and Coors (1988). These comparisons, along with much larger proportions of dominance variance estimated for yield in several maize breeding populations (Edwards and Lamkey 2002; Wardyn *et al.* 2007) suggest that intensive selection and coincident bottlenecking due to maize breeding in those populations depleted additive genetic variation, leading to relatively higher proportions of dominance variance while decreasing inbreeding depression at the same time.

Negative covariances between additive and homozygous dominance deviations reduce the variance among selfed families

We can predict the expected genetic variance among selfed families based on the additive and dominance variances estimated from outbred individuals, ignoring higher-order genetic covariance components (Cockerham and Weir 1984). We observed less than 90% of the expected genetic variance among selfed families for all but three traits in maize and for half of the traits in teosinte (Figure S5; Table S3). Additive-by-additive epistatic variance (V_{AA}) cannot account for this difference, since V_{AA} is expected to contribute more to variance among selfed than outbred families (Cockerham and Weir 1984). However, epistatic variances involving dominance contribute less to the variance among selfed families than to the outbred progenies. Ignoring these dominance epistatic variances could result in overestimation of the additive and dominance variances in the outcrossed progenies and, consequently, overprediction of the selfed family variance. In addition to ignoring epistatic variances, we also ignored

higher-order single-locus genetic covariance components that contribute to genetic variation under inbreeding. One of these single-locus components represents the covariance between additive and homozygous dominance deviations (D_1) that, if negative, can contribute to reduced variance among the selfed families (Cockerham and Weir 1984). It is possible for this covariance component to be negative while also having a strongly positive correlation of breeding values measured in selfed versus outcrossed progenies, as mostly observed here. Our results suggest that this covariance parameter is negative for most traits in maize, but less commonly negative in teosinte.

Negative estimates of this covariance parameter have been reported for yield or fitness-related traits in maize breeding populations and in wild species (Coors 1988; Edwards and Lamkey 2002; Kelly and Arathi 2003; Wardyn *et al.* 2007). For levels of dominance between 0 and *a* (equal to the additive gene action, i.e., complete dominance), this component can be only negative when recessive alleles contributing to inbreeding depression are predominantly the major alleles (Figure S6; Cornelius 1988; Lynch and Walsh 1998; Kelly 1999; Kelly and Willis 2001), which seems unlikely for wild and domesticated landrace populations, and even more unlikely for improved breeding populations. However, for overdominance, there is a range of minor allele frequencies for the unfavorable allele that will generate this negative value (Figure S6). Although true overdominance is rare in maize, pseudo-overdominance generated by repulsion phase linkages among deleterious variants appears to be widespread (Graham *et al.* 1997; Garcia *et al.* 2008; Larièpe *et al.* 2012; Yang *et al.* 2017a), and could produce the negative covariances between additive and homozygous dominance deviations observed in this study. In addition, it is possible that genetic variance is not independent of the mean in general, such that the less than expected amount of genotypic variance observed is due to a reduction in scale of most traits under inbreeding (although we still observe less S_1 variance than expected for flowering time traits in maize that have a higher mean value under inbreeding).

Rare allele effects detected by linkage scan

The rare allele linkage scan is optimized to detect the effect of a private allele carried by a single parent with a contrasting effect to the common alleles carried by other parents. Thus, this linkage scan estimates local effects of each haplotype carried by each parent (Figure 4; Figure S1; Table S8). More than twice as many rare allele QTL were detected in teosinte (468) than maize (173; Figures S7, S8, and S9, Table S8). The relative proportion of QTL detected for different trait groups was nearly identical within maize (29 – 36% of QTL in each group), but within teosinte, most QTL were detected for reproductive traits (58%, compared to 16% for environmental response and 26% for vegetative/flowering time traits). Numerous QTL affecting multiple traits were detected (Tables S9 and S10), reflecting pleiotropic or linked QTL. For example, a QTL allele on chromosome 8 carried by maize parent 164_8 affected six reproductive traits: cupules per row (CUPR), ear diameter (ED), grains per ear (GE), ear internode length (EILN), total number of grains per plant (TGPP), and weight per grain (GW; Figure 4); the rare founder allele reduced the ear length (EL) and consequently reduced the number of cupules per row (CUPR), grains per ear (GE), and total grains per plant (TGPP), while increasing internode length (EILN) and weight of the fewer kernels produced per plant (GW; Table S9). The plants

inheriting the rare allele from teosinte founder PC_O51_ID2 at a QTL on chromosome 6 were shorter and later-flowering, having more but smaller ears, and produced fewer but heavier seeds (Table S10).

The mean size of 2-LOD support interval across all QTL was 38 cM for maize and 26 cM for teosinte, indicating generally low resolution of the parent-specific rare allele QTL, due to the reliance of the RAS on within-family linkage. There was no evidence for QTL tending to occur in low-recombination regions, but the low resolution of QTL positions precludes making strong inferences from this result.

To check how the sample size of progenies (which varies among parents) influences the rare allele scan results, we estimated the correlation between the number of QTL detected and the number of progenies per parent, observing moderate correlations within each population (r = 0.38, p = 0.04 in maize, r = 0.45, p = 8.3 x10-04; Figure S10). Nevertheless, more total QTL and more QTL per parent (10 QTL per parent in teosinte vs 5 per parent in maize on average) were detected in teosinte despite fewer offspring per parent (231 on average in maize compared to 182 on average in teosinte). Therefore, differences in family size between maize and teosinte do not appear to account for the larger number of QTL detected in teosinte.

Relatively few traits exhibited evidence for shared QTL positions between maize and teosinte, suggesting that most private allele variants detected with the RAS were unique to each population (Figure S11). However, we found significant non-random overlaps between RAS QTL intervals and individual SNPs previously detected by Chen et al. (2020) with GWAS for most traits in teosinte and some traits in maize (Figures S11, S12, S13, S14, S15, S16, and S17; Tables S11 and S12). This suggests that there is some common part of the genetic architecture identified by both methods. Further evidence of this are the significant correlations between the standardized effects of rare allele QTL and minor alleles detected by GWAS within common QTL intervals (r = 0.49 and 0.60 for additive and dominance effects in maize; r = 0.32 and 0.57 for additive and dominance effects in teosinte, all $p <$ 0.0001; Figure S18). In particular, Chen et al. (2020) demonstrated that most large-effect QTL had low minor allele frequency, presumably due to selection against large effect variants; the largest of those large-effect SNP associations were also detected in our rare allele scan as QTL, for example the major plant height QTL on chromosome 7 in maize and the major grain weight QTL on chromosome 4 in teosinte (Figure 4).

The additive and dominance effects of RAS QTL were also strongly correlated (r = 0.83 and 0.73 in maize and teosinte, respectively, $p <$ 0.0001; Figure S19), congruent with GWAS results (Chen *et al.* 2020). Thus, the preponderance of QTL effects detected either by GWAS or by RAS had dominance effects in the same direction as the major allele and also were consistent with the direction of observed inbreeding depression for the trait (Figure S20). In contrast to the general trends of positive major allele effects on reproductive traits, we nevertheless observed one GWAS association each for TGWP in maize and GW in teosinte and two RAS QTL for ED in teosinte with standardized major allele additive effects less than -1.5, representing rare favorable alleles for these traits; the dominance effects in all of these cases were also negative, indicating that the rare favorable alleles were also recessive. These QTL cause effects counter to genome-wide trends for inbreeding depression.

Rare QTL allele effects are associated with a minority of genetic variation and inbreeding depression

Although we detected a large number of rare allele QTL effects, the total proportion of trait variation that can be attributed to these QTL is low for all traits (Figure 5). Most of the variation is due to polygenic background effects (modeled here using principal components of the marker data), with rare-allele scan QTL contributing an average of 26% (range 9% to 54%) of the total genetic variance in maize and 30% (range 23% to 41%) in teosinte. Within both populations, the mean proportion of genetic variance due to rare allele scan QTL was lowest for reproductive traits (24% to 25%) and largest for environmental response traits (33% in maize and 35% in teosinte). This may reflect stronger selection pressure against large effect variants affecting reproductive traits due to their higher correlation with fitness.

In general, the amount of observed inbreeding depression that could be accounted for by RAS QTL was below 20%, and the predicted direction of inbreeding depression was wrong in a handful of cases (Figure S21). The best predictions were made for EILN in teosinte and TILN in maize, for which between 40 and 50% of the observed inbreeding depression was predicted by the RAS QTL. GWAS associations predicted a bit more of the observed inbreeding depression, but again this was a relatively small proportion of the total inbreeding depression for most traits. In general, the predicted inbreeding depression was greater for teosinte than maize, probably due mostly to the much larger number of QTL detected in teosinte. In summary, our results indicate that we can detect rare variants contributing to inbreeding depression, but most variation for inbreeding depression appears to be due to polygenic small effect variants that are difficult to detect with either GWAS or RAS, are not easily purged by natural selection, and have increased in importance in maize compared to teosinte.

Early seedling lethal variants inferred from segregation distortion

We sampled leaf tissue about one month after sowing; during this time 15% of maize plants and 49% of teosinte plants did not germinate or survive the first month to be sampled for leaf tissue. Of the plants that survived the first month, an additional 10% of maize plants and 3% of teosinte plants died before maturity. Strongly deleterious alleles, therefore, may have been purged from our self-fertilized families before they could contribute to inbreeding depression of the traits we measured on adult plants. We should be able to detect such loci by testing for segregation distortion at each marker in the families we were able to genotype. We caution, however, that although we removed the seed coat, teosinte varies for seed dormancy, a trait adaptive in natural populations (Avendaño López *et al.* 2011), and it is possible that some seeds that did not germinate may have been alive but dormant. We identified 3 significant segregation distortion regions (SDR) across three different chromosomes in 3 self-families in maize landraces and 16 significant SDR on seven different chromosomes in 11 self-families in teosinte (Figure 6; Table S13). We checked for correspondence between SDR and six QTL for germination rate in teosinte previously identified by Chen et al. (2020): three of the six QTL mapped to the same chromosomes as SDR, one of them inside an SDR (chromosome 1) and one near an SDR (chromosome 5;

Figure S24). These last two cases may represent SDR caused by seed dormancy rather than by selection against lethal recessives, leaving 14 SDR not closely linked to a seed dormancy QTL.

Segregation distortion was mostly observed against either one or two homozygous classes in both populations (Figure S22), consistent with the expectation that lethal or strongly deleterious alleles are recessive and maintained only in heterozygotes. Some SDRs exhibited a deficit of both homozygous classes, consistent with multiple deleterious recessive variants carried in repulsion phase linkage in the parent of the affected family. An exception to this expected pattern was observed in five of the 16 SDRs identified in teosinte, where the heterozygote class was deficient (Figure S22). Some of these cases could represent deleterious alleles that are not recessive (such as the SDR allele on chromosome 7 carried by parent PC_I05_ID1 which is completely absent in its selfed progenies). We also noted that three of the SDR detected in teosinte corresponded with the three best characterized gametophyte factor loci (*Ga1*, *Ga2*, and *Tcb1*) known to control preferential fertilization (Lu *et al.* 2020) and to exist in teosinte (Kermicle *et al.* 2006; Kermicle and Evans 2010). These loci may contribute to reproductive isolation between populations and could be polymorphic in some populations due to complex balancing selection (Kermicle *et al.* 2006; Kermicle 2006). Self-fertilization of individuals heterozygous at these loci is expected to result in reduced frequency of one allele due to selection against pollen carrying one of the two alleles, with no selection occurring on the female gametes. The observed patterns of segregation distortion at these loci did not match this pattern (Figure S25), however, suggesting that they are not the direct result of gametophyte factors, but instead due to accumulation of deleterious alleles linked to 'selfish' gametophyte alleles.

All but five of the 19 SDRs across both populations were detected in regions with lower recombination rates than their respective chromosome average (Table S13 and Figure S23). The observed correspondence between low recombination rates and regions carrying strongly deleterious alleles that are eliminated before the adult life stage is consistent with the prediction that selection is less effective in lower recombination regions (Hill and Robertson 1966; Waller 2021) and previous observations in maize inbred lines (McMullen *et al.* 2009). This result contrasts with the lack of a relationship between recombination rate and RAS QTL, perhaps because the individual QTL effects on fitness are smaller than the SDR effects, and these QTL are under weaker selection.

Parental rare allele load is not consistently related to inbreeding depression

We compared the genome-wide rare allele load of each parent to its percent inbreeding depression for each trait, measured based on its selfed progenies. No relationship was observed for most traits, but significant ($P < 0.05$) positive correlations between the proportion of loci carrying rare alleles in a parent and the inbreeding depression measured in its offspring was observed for days to silk (DTS) and GW in maize ($r = 0.37$ and 0.41, respectively; Figure 7; Table S14). Rare allele load was negatively correlated with inbreeding depression for CUPR and EILN in maize ($r = -0.37$, $r = -0.36$, respectively), indicating that parents with more rare alleles overall had less difference between their outbred and inbred progeny, counter to the hypothesis that rare alleles tend to contribute to inbreeding depression for these traits.

We also measured relationships between inbreeding depression and proportions of rare alleles within RAS QTL intervals for the tested trait. In this case, we found significant positive correlations only for GW in maize (r = 0.35, p = 0.048) and total grain weight per plant (TGWP) in teosinte (r =0.30, p = 0.037; Figure 7; Table S15). Thus, we found a relationship between rare allele load within QTL intervals and parent-specific inbreeding depression only for one reproductive trait in teosinte only, and the correlation value was not very strong. Our inbreeding depression QTL are detected with limited genomic resolution, such that the QTL intervals used to define regions in which to measure rare allele load probably contain too many neutral variants for us to detect any relationships.

Summary

The domestication bottleneck has resulted in reduced genetic diversity and a higher load of segregating predicted deleterious mutations in the global maize gene pool compared to teosinte (Hufford *et al.* 2012a; Wang *et al.* 2017; Lozano *et al.* 2021). Consistent with the reduced diversity of worldwide maize, we find that this particular pair of teosinte and maize populations shares 4.2 M segregating SNPs, with the maize population exhibiting 84.8% of the nucleotide diversity found in teosinte (Chen *et al.* 2020). The site frequency spectra of the two populations reveal that maize is enriched for intermediate frequency derived alleles compared to teosinte (Figure 2). However, we detected fewer rare allele scan QTL and regions of significant segregation distortion in maize than in teosinte. The higher mean GERP score per site observed in teosinte presents an explanation: teosinte has more mutations of larger effect with lower allele frequencies, which would result in detecting more rare allele scan QTL in teosinte. In contrast, it appears that maize carries more common, smaller-effect mutations that cumulatively result in more inbreeding depression, of which a lower proportion can be attributed to detectable QTL.

The maize landrace studied here has a higher segregating genetic load predicted from evolutionary conservation of sequence variants, a higher frequency of derived alleles, and suffers more from inbreeding depression than its counterpart wild teosinte population collected from a nearby location. The genetic architecture of inbreeding depression in both maize and teosinte appears to be dominated by polygenic, small-effect recessive variants that individually are subjected to only limited selection pressure, but cumulatively contribute to substantial inbreeding depression (up to about 50% of seed yield in the maize population). In addition, rare larger-effect variants distributed among the parents of our study sample contribute a smaller but still substantial amount of genetic variation, as documented here as rare allele QTL effects, and these are more important in teosinte than in maize. Finally, an even rarer group of variants appears to be associated with juvenile lethality and are more important in teosinte than in maize. The more polygenic architecture of the segregating genetic load of our maize population renders it more recalcitrant to purging selection and poses a hindrance to modern breeding with landrace populations.

Methods

Population sampling, field evaluation, genotyping, and parentage inference

A sample of 70 teosinte plants was collected from a single field from the near the town of Palmar Chico in the Balsas river drainage of Mexico and 55 maize plants (classified as landrace Tuxpeño) plants were sampled from a single field nearby. Teosinte plants were selfed and intermated; 49 of these parental plants produced sufficient seed for progeny family evaluations (Tables S1 and S2). Similarly, maize landrace plants were selfed and crossed, and 40 of the parental plants produced sufficient seed for progeny evaluations (Tables S1 and S2).

Progenies were evaluated over two winter seasons under short daylengths (<12 h) in Homestead, Florida; details of the field experimental methods are given in Yang et al. (2019). Separate but nearby field blocks were used for teosinte and maize evaluations. Progenies were planted in randomized positions in the field blocks; at planting only the female parents of the progenies were known. Progenies and parents were genotyped using genotyping-by-sequencing (GBS) as described in Yang et al. (2019). The parentage of each progeny was determined using the GBS data of the parents and progeny plants, and this allowed us to distinguish selfed from outcrossed progeny. Eighteen traits were scored on both the teosinte and landrace progeny, these data were previously reported in Yang et al. (Table 1; 2019).

GBS data were filtered for missing call rates, minor allele frequency, and imputed to obtain a total of 34,899 SNPs called for teosinte and 40,255 SNPs for maize landrace. The imputation method accounted for the parentage of each individual and the relatively high error rates for calling heterozygous sites that can occur with low coverage sequence data, as detailed in Yang et al. (2019), and as implemented in ParentPhasingPlugin and ImputeProgenyStatesPlugin in TASSEL5 (https://bitbucket.org/tasseladmin/tassel-5-source/src/master/src/net/maizegenetics/analysis/imputation/; Bradbury *et al.* 2007). Briefly, the imputation involves phasing the parents using progeny information and inferring the parent state at each site using a hidden Markov model. In addition to the SNP calls at each site, therefore, this imputation method calls the most likely parental haplotype inherited by each progeny at each site, using an arbitrary coding for parental haplotypes that is consistent across all selfed and outcrossed progenies of a particular parent. Following quality control of trait and SNP data, we retained data on 4,455 teosinte progeny and 4,398 maize progeny.

Quantitative genetic analysis of inbreeding depression: Diallel model

We fitted a sequence of four diallel linear mixed models to the data for each trait and population (Möhring *et al.* 2011; Isik *et al.* 2017). The simplest model (model 1) was:

$Y_{ijkl} = E_i + T_j + ET_{ij} + x_{sij}\beta_S + x_{Bij}B(E)_i + x_{Rij}\beta_{R1i} + x_{Rij}{}^2\beta_{R2i} + x_{Rij}{}^3\beta_{R3i} + x_{Rij}{}^4\beta_{R4i} + x_{Cij}\beta_{C1i} + x_{Cij}{}^2\beta_{C2i} + x_{Cij}{}^3\beta_{C3i} + xR_{ij}{}^4\beta_{C4i} + M_{Ojk} + F_{Ojl} + M_{Sjk} + MF_{Ojkl} + ME_{Oijk} + FE_{Ojl} + ME_{Sijk} + MFE_{Oijkl} + \varepsilon_{ijkl}$, where:

Y_{ijkl} is the observed phenotype on individual *j* (with maternal parent *k* and paternal parent *l*) in environment *i*.

Model 1 includes the following fixed effects:

E_i is the effect of environment (year) *i*

T_j is the effect of progeny type (selfed vs outcrossed) of individual *j*

ET_{ij} is the interaction of progeny type and environment

x_{sij} is the deviation of the shading measurement on the *ij*th individual from the overall average shading measurement, described in Yang et al. (2019)

β_S is the effect of shading of neighboring plants

x_{Bij} is a dummy variable indicating that individual *j* in year *i* is adjacent to a border row or a tractor tire track

$B(E)_i$ is the effect of border or tractor tire adjacent rows within year *i*

x_{Rij} and $x_{Cij}{}^p$ are *p*= first to fourth order polynomials of the deviation in the row and column directions, respectively, of the *ij*th plant's position from the center of the field in year *i*,

β_{Rpi} and β_{Cpi} are the regression coefficients associated with the *p*th polynomials for row and column trend effects within year *i*, respectively.

Model 1 also includes the following random effects:

M_{Ojk} is the effect of the mother *k* of progeny *j* if *j* is outcrossed; $M_{Ojk} \sim N(0, A_k\sigma^2_{GCA})$, where $\boldsymbol{A_k}$ is the matrix of pedigree relationships among the parents (some of the teosinte parents were themselves half-sibs), and σ^2_{GCA} is the general combining ability (GCA) variance of the parents.

F_{Ojl} is the effect of the father *i* of progeny *j* if *j* is outcrossed. The effect of a particular father is constrained to be equal to the effect of the same parent as a mother, so the distribution of father effects is equal to the mother GCA effects: $F_{Ojl} = M_{Ojk}$ if $k = i$.

M_{Sjk} is the effect of the mother *k* of progeny *j* if *j* is selfed; $M_{Sj} \sim N(0, A_k\sigma^2_{S1})$, where σ^2_{S1} is the variance among selfed (S_1) families.

MF_{Ojkl} is the interaction effect of the mother *k* and father **l** of progeny *j* if *j* is outcrossed; $MF_{Ojkl} \sim N(0, \sigma^2_{SCA})$, where σ^2_{SCA} is the specific combining ability (SCA) variation.

ME_{Oijk}, FE_{Oijl}, ME_{Sijk}, and MFE_{Oijkl} are interactions between environment (year) and the previously defined mother and father effects. ME_{Oijk} and $FE_{Oijk} \sim N(0, A_k \otimes I\sigma^2_{GCA*E})$, where $A_k \otimes I$ is Kronecker product of the matrix of pedigree relationships among the parents (A_k) and **I** is a 2 × 2 identity matrix, and σ^2_{GCA*E} is the GCA-by-environment interaction variance. $ME_{Sijk} \sim N(0, A_k \otimes I\sigma^2_{S1*E})$, where σ^2_{S1*E} is the S_1 family-by-environment interaction variance. $MFE_{Oijkl} \sim N(0, \sigma^2_{SCA*E})$, where σ^2_{SCA*E} is SCA-by-environment variation.

ε_{ijkl} is the residual effect; $\varepsilon_{ijkl} \sim N(0, \sigma^2_{\varepsilon i})$, where $\sigma^2_{\varepsilon i}$ is the residual variance for year *i* (separate residual variances were fit for each year).

Model 2 includes all the same effects as model 1, but separates the residual variance for outcrossed and selfed progenies, so that four distinct residual variances were estimated: $\sigma^2_{\varepsilon O1}$ is the residual variance for outcross progeny in the first year, $\sigma^2_{\varepsilon O2}$ is the residual variance for outcross progeny in the second year, $\sigma^2_{\varepsilon S1}$ is the residual variance for selfed progeny in the first year, and $\sigma^2_{\varepsilon S2}$ is the residual variance for selfed progeny in the second year.

Model 3 includes all the same effects as model 2 and adds a covariance between the effects of a parent on outcrossed and selfed progenies:

M_{Ojk} and M_{Sjk} are jointly distributed as $N(0, \boldsymbol{A} \otimes \sum)$, where $\sum = \begin{bmatrix} \sigma^2_{GCA} & \hat{\sigma}_{GCA,S1} \\ \hat{\sigma}_{GCA,S1} & \sigma^2_{S1} \end{bmatrix}$ and $\hat{\sigma}_{GCA,S1}$ is the covariance between parent effects on outcrossed progeny (GCA) and selfed progeny.

Model 4 includes all the same effects as model 3 but adds reciprocal parental genetic effects:

M_{ORjk} is the reciprocal effect of the mother k of progeny *j* if *j* is outcrossed, it measures the difference between the effect of parent k when used as female vs male. $M_{ORjk} \sim N(0, \sigma^2_{RGCA})$, where σ^2_{RGCA} is the reciprocal general combining ability (GCA) variance of the parents.

F_{ORjl} is the reciprocal effect of the father *l* of progeny *j* if *j* is outcrossed.

MF_{ORjkl} is the reciprocal effect that distinguishes the interaction of the mother *k* and father *l* progeny *j* vs that interaction effect for the reciprocal cross if *j* is outcrossed; $MF_{ORjkl} \sim N(0, \sigma^2_{RSCA})$, where σ^2_{RSCA} is the reciprocal SCA variation.

MFE_{ORjkl} is the interaction between environment (year) and the reciprocal SCA effect. $MFE_{ORjkl} \sim N(0, \sigma^2_{RSCA*E})$, σ^2_{RSCA*E} is the reciprocal SCA-by-environment variation.

For every trait and population, all four models were fit using ASReml version 4.2 (Gilmour *et al.* 2021). To obtain convergence with models 3 and 4, the covariance between GCA and S1 progeny effects was unconstrained, such that resulting estimates could correspond to correlations outside of the bounds of -1 to + 1. Models were compared using Bayesian Information Criteria (BIC). From selected models, the following parameters were estimated (Hallauer and Miranda Filho 1988; Falconer and Mackay 1996):

$$\hat{\sigma}^2_A = 4\hat{\sigma}^2_{GCA}$$

$$\hat{\sigma}^2_D = 4\hat{\sigma}^2_{SCA}$$

$h^2 = \frac{\hat{\sigma}_A^2}{0.5\hat{\sigma}_A^2 + 0.25\hat{\sigma}_D^2 + 2\hat{\sigma}_{GCA*E}^2 + \hat{\sigma}_{SCA*E}^2 + \overline{\hat{\sigma}_{\varepsilon 0}^2}}$, where $\overline{\hat{\sigma}_{\varepsilon 0}^2}$ is the average residual variance for outcross progenies from each year. The residual variance includes both within-full-sib-family genetic variance ($0.5\hat{\sigma}_A^2 + 0.75\hat{\sigma}_D^2$) and experimental error variance.

The marginal mean values of each progeny type (outcrossed and selfed) were predicted at common values of the fixed effects. The coefficient of inbreeding depression was estimated as the proportional change in population mean between outcrossed and selfed progenies: $\delta = 1 - \frac{outcrossed\ mean}{\text{selfed mean}}$ (Roff 1997).

The genetic variance among selfed ($S_{0:1}$) families is expected to be $\sigma_A^2 + 0.25\sigma_D^2 + D_1 + 0.125D_2$, ignoring epistatic variances, where D_1 is the covariance between additive and homozygous dominance deviations and D_2 is the variance of homozygous dominance deviations (Cockerham 1983). Assuming that D_1 and D_2 are negligible, we can predict the selfed family variance to be $\sigma_A^2 + 0.25\sigma_D^2$, using estimates of additive and dominance variance components from the outbred progenies. If the ratio of observed selfed family variance to this predicted value is greater than one, it is evidence for the importance of epistatic or D_2 variances, or a positive D_1 covariance. In contrast, if the ratio of the observed to predicted selfed family variance is less than one, it is evidence for a negative D_1 covariance.

Regression of trait values on genomic inbreeding coefficient estimate

The genomic realized estimate of the inbreeding coefficient (*F*) of each individual was obtained from the realized additive genomic relationship matrix as 1 minus the diagonal elements (Endelman and Jannink 2012). This estimate of the inbreeding coefficient was included in a separate linear mixed model that contained the same fixed covariates (except for $T_j + ET_{ij}$) and the same residual effects distribution as diallel model 1 and adding the genomic estimated inbreeding coefficient and its interaction with environments as fixed covariates. In addition, the model included these random effects:

Aij, the polygenic additive effect of the *ij*th plant, with distribution $A_{ij} \sim N(0, A\sigma_A^2)$, where **A** is the realized additive genomic relationship matrix and σ_A^2 is the additive genetic variance,

D_{ij}, is the polygenic dominance effect of the *ij*th plant, with distribution $D_{ij} \sim N(0, D\sigma_D^2)$, where **D** is the realized additive genomic relationship matrix and σ_D^2 is the dominance genetic variance,

$G{\times}E_{ij}$ is the interaction of polygenic effect of the *j*th plant in environment *i*, with distribution $G{\times}E_{ij} \sim N(0, A_1 \oplus A_2\sigma_{AE}^2)$, where $\boldsymbol{A_t}$ is the realized additive relationship matrix for individuals grown in year *i* and $(A_1 \oplus A_2)$ is a block-diagonal structure that includes non-zero covariances between plants grown in the same year, but zero covariance between plants grown in different years.

This model was fit with ASReml version 4.2, and the regression effect of the inbreeding coefficient was tested for significance using a conditional F-test. The slope of regression on the inbreeding coefficient was estimated as its main effect plus the average of its interactions with years.

The influence of epistasis on inbreeding depression can be tested by including the squared inbreeding coefficient as an additional regression variable (Falconer and Mackay 1996; Lynch and Walsh 1998). We computed the square of the genomic estimated inbreeding coefficient (fixing the squared value of the slightly negative estimates for some individuals as zero) and fit this coefficient and its interaction with years in a second model otherwise identical to the previous model.

Rare allele scan (RAS) to detect private haplotype effects

To detect large-effect rare variants that are carried by only one parent, we developed a novel linkage scan that tests each genetic map position for the effect of one parental allele compared to all other parental alleles (Figure S1). This analysis relies on progeny genotype calls that represent identity-by-descent parental allele calls that were generated during the imputation and phasing of the GBS data described in detail in Yang et al. (2019). These genotype calls reflect the phased parent haplotypes inherited by each individual progeny at each SNP.

The first step of the rare allele scan (RAS) was to create separate linkage maps for each population. Recombinations of parental haplotyes observed in the progeny were used to estimate linkage distances between SNPs. Within each interval of 1 cM we expect to observe that 1% of the progeny haplotypes have recombined compared to their parents. We selected SNP markers at 1 cM intervals to represent the linkage maps by identifying marker pairs separated by 0.01*(2N) recombinations, where N is the genotyped progeny sample size. The resulting linkage maps had total distances of 1371 and 1426 cM for teosinte and maize, respectively.

In a second step, were created a new data set for each parental haplotype, in which the progeny haplotype calls were recoded to the number of copies of one parental haplotype carried by each progeny at each of the linkage map markers (Figure S1). Then, we fit a linear model for each combination of parental haplotype and linkage marker:

$Y_{ij} = E_i + F_{ij}\beta_{Fi} + x_{sij}\beta_S + x_{Bij}B(E)_i + x_{Rij}\beta_{R1i} + x_{Rij}{}^2\beta_{R2i} + x_{Rij}{}^3\beta_{R3i} + x_{Rij}{}^4\beta_{R4i} + x_{Cij}\beta_{C1i} + x_{Cij}{}^2\beta_{C2i} + x_{Cij}{}^3\beta_{C3i} + xR_{ij}{}^4\beta_{C4i} + PC1_{ij} + PC2_{ij} + \ldots + PC50_{ij} + x_{haij}a + x_{hdij}d + \varepsilon_{ij}$, where:

E_i is the effect of environment (year) *i*

F_{ij} is the genomic estimated inbreeding coefficient of individual *j* within year *i*

β_{Fi} is the effect of inbreeding within year *i*

$PC1_{ij} + PC2_{ij} + \ldots + PC50_{ij}$ are the scores for individual *j* within year *i* on each of the first 50 principal components of the genome-wide SNP data, previously computed by Chen et al. (2020)

x_{haij} is the number of the parental haplotype alleles carried individual *j* within year *i* at the tested map position

a is the additive effect of the parental allele at the tested map position

$x_{\mathrm{hd}ij}$ is an indicator for heterozygosity for the current parental haplotype allele in individual *j* within year *i* at the tested map position

d is the dominance effect of the parental allele at the tested map position

E_i, x_{sij}, β_S, x_{Bij}, $\mathrm{B(E)}_i$, $x_{Rij}{}^p$, $x_{Cij}{}^p$, β_{Rpi}, and β_{Cpi} are as defined in the diallel model previously.

Models were fit using ordinary least squares, implemented in the statsmodels package of Python (Seabold and Perktold 2010). The LOD score for each marker-parent haplotype combination was calculated as:

LOD = $(n/2)\log_{10}(\mathrm{RSS0/RSSmark})$, where *n* is the progeny sample size, RSSmark is the residual sum of squares of the full model shown above, and RSS0 is the residual sum of squares from a model all of the effects in the full model except for the marker additive and dominance effects.

This model is equivalent to the genome-wide association test implemented by Chen et al (2020) except that it tests the additive and dominance effects of a single parental allele defined by identity-by-descent instead of the usual additive and dominance effects of SNP alleles defined by identity-in-state (Figure S1).

Initial putative RAS quantitative trait loci (QTL) positions were declared at marker-haplotype combinations with LOD scores greater than or equal to 4.0. LOD-2 support intervals for the initial QTL scans for each parental haplotype were defined based on the results of the initial linkage scan as contiguous regions of the genetic map LOD scores within 2 LOD of the largest effect in the region.

The final step of this analysis was a forward stepwise regression fitting the model above for the most significant marker-haplotype and then adding each initially declared QTL to the model, one at a time. The QTL that improved (decreased) the Bayesian Information Criterion (BIC) of the model most was added to the model. This process was repeated until no more QTL improved the model BIC. Only QTL remaining in the final model were reported as significant QTL. QTL number, effects, individual variances, and combined variances were estimated from this final model.

Testing for non-random clustering of RAS QTL and GWAS association positions

Physical positions of linkage map markers were located on the maize B73 AGP version 4 reference sequence (Jiao *et al.* 2017), as were SNP markers used for GWAS previously by Chen et al. (2020). This allowed estimation of the physical sequence size and local recombination rate (cM per Mbp) for RAS QTL and identification of previously-identified trait-associated SNPs inside of QTL for the same trait. For each trait, the proportion of RAS QTL containing associated SNPs identified by GWAS, the proportion of QTL overlapping within a population (because they were detected in different parents), and the proportion overlapping between populations were computed using the 2-LOD support intervals to define QTL windows. The distributions of the proportion of QTL containing GWAS associations or QTL overlapping within or between populations by chance were simulated by a permutation test. A single permutation of QTL positions was made by randomly assigning windows corresponding to the cM size of the observed QTL support intervals for a trait to the linkage map. The proportion of GWAS associations

within these permuted windows represents one realization of the proportion of overlap between GWAS and QTL positions under the null hypothesis of random QTL positions with respect to GWAS associations. Similarly, the proportion of QTL positions overlapping within a population within one permuted data set represented the proportion of overlapping QTL among parents under the null hypothesis of random distribution of QTL positions among parents. Finally, to estimate the proportion of overlapping QTL between maize and teosinte for a common trait, QTL window positions were permuted within each population and compared. We repeated this process 5000 times for each set of comparisons. This provided an empirical estimate of the distribution of the proportion of overlaps under the null hypothesis of random QTL positions. Two-sided empirical *p*-values for the observed proportion of QTL overlaps were estimated based on the percentile of the permuted distribution corresponding to the observed proportion of overlaps for a given comparison.

Estimation of inbreeding depression associated with rare allele scan QTL and genome-wide associated SNPs

We predicted the amount of population inbreeding depression expected to be caused by a single RAS QTL as 2*Fpqd* (Falconer and Mackay 1996), where *F* is the inbreeding coefficient, and *p* and *q* are the major and minor allele frequencies. W set *F* = 0.5 for the one generation of selfing used in this experiments, and *q* = 1/2N for N parents because we model the RAS QTL alleles as carried on only one haplotype of one parent, and *d* is the estimated dominance effect of the QTL scaled to the outcross progeny mean rather than to the phenotypic standard deviation so that we can compare this to the observed percent inbreeding depression observed in selfed versus outcrossed progeny means computed from the diallel analysis. The predicted amount of inbreeding depression associated with each RAS QTL was summed over QTL for a given trait and population and compared to the observed inbreeding depression. A similar comparison was made for GWAS associations, using the minor allele frequency at each QTL observed in the parents as *q*.

Linkage Scan for Segregation Distortion

For both maize LR and teosinte, we used the same consensus linkage maps used in the private haplotype test described above to identify genomic regions exhibiting segregation distortion. Selfed families with 20 or more progenies were included in this analysis. We analyzed 24 and 42 selfed families for maize landraces and for teosinte, respectively. The number of progenies per selfed family analyzed ranged from 20 to 125 (mean 52.9) in maize and from 20 to 95 (mean 45.2) in teosinte. In total, 1,271 selfed progenies were included in this analysis in maize and 1,899 in teosinte. Chi-square tests of goodness of fit of the observed segregation ratios at each linkage map marker to the expected 1:2:1 proportions were performed with the R package OneMap (Margarido *et al.* 2007).

Multiple test correction for the purpose of declaring genome-wide significance thresholds is challenging in genetic studies because of the complex correlation structure among marker tests. We

inferred the effective number of independent tests (ENT) based on the pairwise LD between SNPs (Gao *et al.* 2008). Matrices of correlation coefficients between each pair of SNPs in the consensus linkage map of each chromosome were used to compute ENT for each chromosome using the R package poolR (Cinar and Viechtbauer 2021). The chromosome-specific ENTs were summed to obtain a genome-wide ENT, and the significance threshold for segregation distortion was set as $0.05/ENT_{total}$. Significance thresholds were computed separately for maize and teosinte. Furthermore, we required a contiguous group of at least three consecutive markers passing the whole genome threshold to declare a significant segregation distortion region (SDR).

Finally, to ensure that putative significant SDR were not an artefact of the imputation and phasing procedure, we performed a check on the segregation ratios of genotype calls in the raw GBS data. The first step was to delimit windows corresponding to each SDR. In a second step, we selected the raw GBS markers that passed the quality control filters imposed before imputation described by Yang et al. (2019) within the SDR windows for the specific parent of the family where segregation distortion was observed and that were heterozygous in the parent. Then we obtained the P-values for the chi-square tests for goodness of fit to the expected 1:2:1 segregation ratio in the progenies of the affected family. We required more than half of the informative raw GBS markers to have significant P-values ($P < 0.05$) for these chi-square tests to confirm the SDR originally identified with the imputed data. We calculated the recombination rate in each significant SDR as cM length of the region divided by the physical length of the region in Mbp. For comparison, we computed both the genome-wide average recombination rate per Mbp and also the average recombination rate specific to the chromosome on which the SDR was located.

The relationship between parent-specific rare allele load and inbreeding depression

Genotypes at 20,631,107 and 17,819,152 SNPs were obtained for each parent of the maize and teosinte, populations, respectively (Chen *et al.* 2020). To compute the genome-wide rare allele load of each parent, we first subset the complete set of SNPs to include only those with minor allele frequencies < 0.05 (considered rare in our study). After this filter 9,444,274 and 9,823,305 markers were retained for maize and teosinte, respectively. Each parent's rare allele load was the mean frequency of rare alleles across these loci. We computed the correlation between each parent's rare allele load and parent-specific percent inbreeding depression for each trait within each population. We included both homozygous and heterozygous rare allele calls in this computation (with homozygous calls given double weight) because inbreeding depression was measured as the difference between each parent's outbred and inbred progeny values, not as the difference between the parent itself and its selfed progeny. Thus, although homozygous recessive deleterious alleles in the parent do not contribute to a difference between the parent and its inbred progeny, they do contribute to a difference between inbred progeny and the outbred progenies of that parent (as they will be almost entirely complemented by common alleles from other parents in outbred progenies). Next, we computed the rare allele load carried by each parent at the subset of genomic positions within the support intervals of all rare allele scan QTL for a

given trait. For each trait, correlations were estimated between the proportion of inbreeding depression exhibited by each parent and the parent-specific QTL rare allele load.

Estimation of mutational burden per parent

Genetic load for each parent was estimated from genomic evolutionary rate profiling (GERP) scores using GERP scores computed by Kistler et al. (2018) from whole genome sequence data. Only sites segregating in the respective population (maize or teosinte) with an allele call in sorghum as outgroup, and a positive GERP score were included. Using the sorghum allele as ancestral allele, per individual burden was calculated under a partially recessive model with a per site burden equal to the GERP score for that site for homozygous derived genotypes, 0.25 × GERP for heterozygous genotypes and 0 for homozygous ancestral genotypes. From this, the total and mean GERP effects across loci were computed per individual.

Acknowledgements

This work was supported by United States National Science Foundation grant IOS 1238014 to E.SB., J.F.D., J.R.I., and Q.S. (https://www.nsf.gov/index.jsp). Q.C. was supported by United States National Science Foundation grant IOS 1934865 to J.F.D. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank Jason Brewer for his assistance in this project.

Data Availability Statement

Data, analysis scripts, and supplemental table files are available at https://figshare.com/s/3dff12b178607a9a1285.

Literature Cited

Avendaño López A. N., J. de Jesús Sánchez González, J. A. Ruíz Corral, L. De La Cruz Larios, F. Santacruz-Ruvalcaba, *et al.*, 2011 Seed Dormancy in Mexican Teosinte. Crop Sci. 51: 2056. https://doi.org/10.2135/cropsci2010.09.0538

Becher H., B. C. Jackson, and B. Charlesworth, 2020 Patterns of Genetic Variability in Genomic Regions with Low Rates of Recombination. Curr. Biol. 30: 94-100.e3. https://doi.org/10.1016/j.cub.2019.10.047

Beissinger T. M., L. Wang, K. Crosby, A. Durvasula, M. B. Hufford, *et al.*, 2016 Recent demography drives changes in linked selection across the maize genome. Nat. Plants 2: 1–7. https://doi.org/10.1038/NPLANTS.2016.84

Bellon M. R., and S. B. Brush, 1994 Keepers of maize in Chiapas, Mexico. Econ. Bot. 48: 196–209. https://doi.org/10.1007/BF02908218

Bellon M. R., A. Mastretta-Yanes, A. Ponce-Mendoza, D. Ortiz-Santamaría, O. Oliveros-Galindo, *et al.*, 2018 Evolutionary and food supply implications of ongoing maize domestication by Mexican campesinos. Proc. R. Soc. B Biol. Sci. 285: 20181049. https://doi.org/10.1098/rspb.2018.1049

Bradbury P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, *et al.*, 2007 TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Brown K. E., and J. K. Kelly, 2020 Severe inbreeding depression is predicted by the “rare allele load” in Mimulus guttatus. Evolution (N. Y). 74: 587–596. https://doi.org/10.1111/evo.13876

Charlesworth D., and B. Charlesworth, 1987 Inbreeding depression and its evolutionary consequences. Annu. Rev. Ecol. Syst. 18: 237–268. https://doi.org/10.1146/annurev.es.18.110187.001321

Charlesworth B., and D. Charlesworth, 1999 The genetic basis of inbreeding depression. Genet. Res. 74: 329–340. https://doi.org/10.1017/S0016672399004152

Charlesworth D., and J. H. Willis, 2009 The genetics of inbreeding depression. Nat. Rev. Genet. 10: 783–796.

Chen Q., L. F. Samayoa, C. J. Yang, P. J. Bradbury, B. A. Olukolu, *et al.*, 2020 The genetic architecture of the maize progenitor, teosinte, and how it was altered during maize domestication. PLoS Genet. 16: e1008791. https://doi.org/10.1371/journal.pgen.1008791

Cinar O., and W. Viechtbauer, 2021 poolr: Methods for pooling the results from (dependent) tests. https://cran.r-project.org/package=poolr

Cockerham C. C., 1983 Covariances of Relatives from Self-Fertilization. Crop Sci. 23: 1177–1180. https://doi.org/10.2135/cropsci1983.0011183x002300060035x

Cockerham C. C., and B. S. Weir, 1984 Covariances of relatives stemming from a population undergoing mixed self and random mating. Biometrics 40: 157. https://doi.org/10.2307/2530754

Coors J. G., 1988 Response to four cycles of combined half-sib and S1 family selection in maize. Crop Sci. 28: 891–896. https://doi.org/10.2135/cropsci1988.0011183x002800060003x

Cornelius P. L., and J. W. Dudley, 1974 Effects of inbreeding by selfing and full-sib mating in a maize population. Crop Sci. 14: 815–819. https://doi.org/10.2135/cropsci1974.0011183x001400060011x

Cornelius P. L., 1988 Properties of components of covariance of inbred relatives and their estimates in a maize population. Theor. Appl. Genet. 75: 701–711. https://doi.org/10.1007/BF00265590

Crow J., and M. Kimura, 1970 *An Introduction to Population Genetics Theory*. The Blackburn Press, Caldwell, New Jersey.

Darwin C. R., 1876 *The Effects of Cross and Self Fertilization in the Veget-able Kingdom*. John Murray, London.

Davydov E. V., D. L. Goode, M. Sirota, G. M. Cooper, A. Sidow, *et al.*, 2010 Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP++, (W. W. Wasserman, Ed.). PLoS Comput. Biol. 6: e1001025. https://doi.org/10.1371/journal.pcbi.1001025

Edwards J. W., and K. R. Lamkey, 2002 Quantitative genetics of inbreeding in a synthetic maize population. Crop Sci. 42: 1094–1104. https://doi.org/10.2135/cropsci2002.1094

Edwards J. W., 2008 Predicted Genetic Gain and Inbreeding Depression with General Inbreeding Levels in Selection Candidates and Offspring. Crop Sci. 48: 2086–2096. https://doi.org/10.2135/cropsci2008.01.0001

Endelman J. B., and J.-L. Jannink, 2012 Shrinkage estimation of the realized relationship matrix. G3 (Bethesda). 2: 1405–13. https://doi.org/10.1534/g3.112.004259

Escobar J. S., A. Nicot, and P. David, 2008 The different sources of variation in inbreeding depression, heterosis and outbreeding depression in a metapopulation of Physa acuta. Genetics 180: 1593–1608. https://doi.org/10.1534/genetics.108.092718

Eyre-Walker A., R. L. Gaut, H. Hilton, D. L. Feldman, and B. S. Gaut, 1998 Investigation of the bottleneck leading to the domestication of maize. Proc. Natl. Acad. Sci. U. S. A. 95: 4441–4446. https://doi.org/10.1073/pnas.95.8.4441

Falconer D. S., and T. F. C. Mackay, 1996 *Introduction to Quantitative Genetics*. Addison Wesley Longman Limited, Edingurgh Gate, Harlow.

Fowler K., and M. C. Whitlock, 1999 The variance in inbreeding depression and the recovery of fitness in bottlenecked populations. Proc. R. Soc. B Biol. Sci. 266: 2061–2066. https://doi.org/10.1098/rspb.1999.0887

Galloway L. F., J. R. Etterson, and J. W. McGlothlin, 2009 Contribution of direct and maternal genetic effects to life-history evolution. New Phytol. 183: 826–838. https://doi.org/10.1111/j.1469-8137.2009.02939.x

Gao X., J. Starmer, and E. R. Martin, 2008 A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. Genet. Epidemiol. 32: 361–369. https://doi.org/10.1002/gepi.20310

Garcia A. A. F., S. Wang, A. E. Melchinger, and Z.-B. Zeng, 2008 Quantitative Trait Loci Mapping and The Genetic Basis of Heterosis in Maize and Rice. Genetics 180: 1707–1724. https://doi.org/10.1534/genetics.107.082867

Gibson G., 2012 Rare and common variants: Twenty arguments. Nat. Rev. Genet. 13: 135–145.

Gilbert K. J., F. Pouyet, L. Excoffier, and S. Peischl, 2020 Transition from Background Selection to Associative Overdominance Promotes Diversity in Regions of Low Recombination. Curr. Biol. 30: 101-107.e3. https://doi.org/10.1016/j.cub.2019.11.063

Gilmour A. R., B. J. Gogel, B. R. Cullis, S. J. Wellham, and R. Thompson, 2021 *ASReml User Guide Release 4.2 Functional Specification*. VSN International Ltd, Hemel Hempstead.

Gonzalo M., T. J. Vyn, J. B. Holland, and L. M. McIntyre, 2007 Mapping reciprocal effects and interactions with plant density stress in Zea mays L. Heredity (Edinb). 99: 14–30. https://doi.org/10.1038/sj.hdy.6800955

Goodman M. M., J. B. Holland, and J. J. Sanchez-Gonzalez, 2014 Breeding and genetic diversity, pp. 41–77 in *Genetics, Genomics and Breeding of Maize*, edited by Wusirika R., Bohn M., Lai J., Kole C. CRC Press, Boca Raton, FL.

Graham G. I., D. W. Wolff, and C. W. Stuber, 1997 Characterization of a yield quantitative trait locus on chromosome five of maize by fine mapping. Crop Sci. 37: 1601–1610. https://doi.org/10.2135/cropsci1997.0011183X003700050033x

Hallauer A. R., and J. H. Sears, 1973 Changes in Quantitative Traits Associated with Inbreeding in a Synthetic Variety of Maize 1. Crop Sci. 13: 327–330. https://doi.org/10.2135/cropsci1973.0011183x001300030012x

Hallauer A. R., and J. B. Miranda Filho, 1988 *Quantitative Genetics in Maize Breeding*. Iowa State University Press, Ames, IA.

Hernández Xolocotzi E., 1985 Maize and man in the Greater Southwest. Econ. Bot. 39: 416–430. https://doi.org/10.1007/BF02858749

Hill W. G., and A. Robertson, 1966 The effect of linkage on limits to artificial selection. Genet. Res. 8: 269–294. https://doi.org/10.1017/S0016672300010156

Hill W. G., M. E. Goddard, and P. M. Visscher, 2008 Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet. 4: e1000008.

Holland J. B., 2009 Increasing Yield, pp. 469–482 in *Handbook of Maize: Its Biology*, edited by Bennetzen J. L., Hake S. C. Springer, New York.

Howard D. M., R. Pong-Wong, P. W. Knap, and J. A. Woolliams, 2017 Use of haplotypes to identify regions harbouring lethal recessive variants in pigs. Genet. Sel. Evol. 49: 57. https://doi.org/10.1186/s12711-017-0332-3

Hufford M. B., P. Gepts, and J. Ross-Ibarra, 2011 Influence of cryptic population structure on observed mating patterns in the wild progenitor of maize (Zea mays ssp. parviglumis). Mol. Ecol. 20: 46–55. https://doi.org/10.1111/j.1365-294X.2010.04924.x

Hufford M. B., X. Xu, J. Van Heerwaarden, T. Pyhäjärvi, J. M. Chia, *et al.*, 2012a Comparative population genomics of maize domestication and improvement. Nat. Genet. 44: 808–811. https://doi.org/10.1038/ng.2309

Hufford M. B., P. Bilinski, T. Pyhäjärvi, and J. Ross-Ibarra, 2012b Teosinte as a model system for population and ecological genomics. Trends Genet. 28: 606–615.

Hufford M. B., P. Lubinksy, T. Pyhäjärvi, M. T. Devengenzo, N. C. Ellstrand, *et al.*, 2013 The Genomic Signature of Crop-Wild Introgression in Maize. PLoS Genet. 9: e1003477. https://doi.org/10.1371/journal.pgen.1003477

Husband B. C., and D. W. Schemske, 1996 Evolution of the timing and magnitude of inbreeding depression in plants. Evolution (N. Y). 50: 54–70. https://doi.org/10.1111/j.1558-5646.1996.tb04472.x

Isik F., J. Holland, and C. Maltecca, 2017 *Genetic Data Analsis for Plant and Animal Breeding*. Springer, New York.

Jiao Y., P. Peluso, J. Shi, T. Liang, M. C. Stitzer, *et al.*, 2017 Improved maize reference genome with single-molecule technologies. Nature 546: 524–527. https://doi.org/10.1038/nature22971

Karczewski K. J., L. C. Francioli, G. Tiao, B. B. Cummings, J. Alföldi, *et al.*, 2020 The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581: 434–443. https://doi.org/10.1038/s41586-020-2308-7

Kelly J. K., 1999 An experimental method for evaluating the contribution of deleterious mutations to quantitative trait variation. Genet. Res. 73: 263–273. https://doi.org/10.1017/S0016672399003766

Kelly J. K., and J. H. Willis, 2001 Deleterious mutations and genetic variation for flower size in Mimulus guttatus. Evolution (N. Y). 55: 937. https://doi.org/10.1554/0014-3820(2001)055[0937:dmagvf]2.0.co;2

Kelly J. K., and H. S. Arathi, 2003 Inbreeding and the genetic variance in floral traits of Mimulus guttatus. Heredity (Edinb). 90: 77–83. https://doi.org/10.1038/sj.hdy.6800181

Kelly J. K., 2005a Family level inbreeding depression and the evolution of plant mating systems. New Phytol. 165: 55–62.

Kelly J. K., 2005b Epistasis in monkeyflowers. Genetics 171: 1917–1931. https://doi.org/10.1534/genetics.105.041525

Kermicle J. L., S. Taba, and M. M. S. Evans, 2006 The Gametophyte-1 locus and reproductive isolation among Zea mays subspecies. Maydica 51: 219–225.

Kermicle J. L., 2006 A selfish gene governing pollen-pistil compatibility confers reproductive isolation between maize relatives. Genetics 172: 499–506. https://doi.org/10.1534/genetics.105.048645

Kermicle J. L., and M. M. S. Evans, 2010 The Zea mays Sexual Compatibility Gene ga2: Naturally Occurring Alleles, Their Distribution, and Role in Reproductive Isolation. J. Hered. 101: 737–749. https://doi.org/10.1093/jhered/esq090

Kirkpatrick M., and P. Jarne, 2000 The effects of a bottleneck on inbreeding depression and the genetic

load. Am. Nat. 155: 154–167. https://doi.org/10.1086/303312

Kistler L., S. Yoshi Maezumi, J. G. De Souza, N. A. S. Przelomska, F. M. Costa, *et al.*, 2018 Multiproxy evidence highlights a complex evolutionary legacy of maize in South America. Science 362: 1309–1313. https://doi.org/10.1126/science.aav0207

Kremling K. A. G., S. Y. Chen, M. H. Su, N. K. Lepak, M. C. Romay, *et al.*, 2018 Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. Nature 555: 520–523. https://doi.org/10.1038/nature25966

Lamkey K. R., and O. S. Smith, 1987 Performance and inbreeding depression of populations representing seven eras of maize breeding. Crop Sci. 27: 695–699. https://doi.org/10.2135/cropsci1987.0011183x002700040017x

Lande R., and D. W. Schemske, 1985 The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models. Evolution (N. Y). 39: 24–40. https://doi.org/10.1111/j.1558-5646.1985.tb04077.x

Larièpe A., B. Mangin, S. Jasson, V. Combes, F. Dumas, *et al.*, 2012 The Genetic Basis of Heterosis: Multiparental Quantitative Trait Loci Mapping Reveals Contrasted Levels of Apparent Overdominance Among Traits of Agronomical Interest in Maize (*Zea mays* L.). Genetics 190: 795–811. https://doi.org/10.1534/genetics.111.133447

Lawson H. A., J. M. Cheverud, and J. B. Wolf, 2013 Genomic imprinting and parent-of-origin effects on complex traits. Nat. Rev. Genet. 14: 609–617.

Lozano R., E. Gazave, J. P. R. dos Santos, M. G. Stetter, R. Valluru, *et al.*, 2021 Comparative evolutionary genetics of deleterious load in sorghum and maize. Nat. Plants 7: 17–24. https://doi.org/10.1038/s41477-020-00834-5

Lu Y., A. N. Moran Lauter, S. Makkena, M. P. Scott, and M. M. S. Evans, 2020 Insights into the molecular control of cross-incompatibility in Zea mays. Plant Reprod. 33: 117–128.

Lynch M., and B. Walsh, 1998 *Genetics and Analysis of Quantitative Traits*. Sinauer, Sunderland, MA.

Margarido G. R. A., A. P. Souze, and A. A. F. Garcia, 2007 OneMap: software for genetic mapping in outcrossing species. Hereditas 144: 78–79. https://doi.org/10.1111/j.2007.0018-0661.02000.x

Matsuoka Y., Y. Vigouroux, M. M. Goodman, J. Sanchez G, E. Buckler, *et al.*, 2002 A single domestication for maize shown by multilocus microsatellite genotyping. Proc. Natl. Acad. Sci. U. S. A. 99: 6080–4. https://doi.org/10.1073/pnas.052125199

McMullen M. D. M., S. Kresovich, H. S. Villeda, P. Bradbury, H. Li, *et al.*, 2009 Genetic properties of the maize nested association mapping population. Science 325: 737–740. https://doi.org/10.1126/science.1174320

Meghji M. R., J. W. Dudley, R. J. Lambert, and G. F. Sprague, 1984 Inbreeding depression, inbred and hybrid grain yields, and other traits of maize genotypes representing three eras. Crop Sci. 24: 545–549. https://doi.org/10.2135/cropsci1984.0011183x002400030028x

Möhring J., A. E. Melchinger, and H. P. Piepho, 2011 REML-based diallel analysis. Crop Sci. 51: 470–478.

https://doi.org/10.2135/cropsci2010.05.0272

Moorad J. A., and M. J. Wade, 2005 A genetic interpretation of the variation in inbreeding depression. Genetics 170: 1373–1384. https://doi.org/10.1534/genetics.104.033373

Paige K. N., 2010 The functional genomics of inbreeding depression: A new approach to an old problem. Bioscience 60: 267–277. https://doi.org/10.1525/bio.2010.60.4.5

Piperno D. R., A. J. Ranere, I. Holst, J. Iriarte, and R. Dickau, 2009 Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. Proc. Natl. Acad. Sci. U. S. A. 106: 5019–5024. https://doi.org/10.1073/pnas.0812525106

Pressoir G., and J. Berthaud, 2004 Patterns of population structure in maize landraces from the Central Valleys of Oaxaca in Mexico. Heredity (Edinb). 92: 88–94. https://doi.org/10.1038/sj.hdy.6800387

Pyhäjärvi T., M. B. Hufford, S. Mezmouk, and J. Ross-Ibarra, 2013 Complex patterns of local adaptation in teosinte. Genome Biol. Evol. 5: 1594–1609. https://doi.org/10.1093/gbe/evt109

Räsänen K., and L. E. B. Kruuk, 2007 Maternal effects and evolution at ecological time-scales. Funct. Ecol. 21: 408–421.

Roessler K., A. Muyle, C. M. Diez, G. R. J. Gaut, A. Bousios, *et al.*, 2019 The genome-wide dynamics of purging during selfing in maize. Nat. Plants 5: 980–990. https://doi.org/10.1038/s41477-019-0508-7

Roff D. A., 1997 *Evolutionary Quantitative Genetics*. Springer US.

Samayoa L. F., J. C. Dunne, R. J. Andres, and J. B. Holland, 2018 Harnessing maize biodiversity, pp. 335–366 in *The Maize Genome. Compendium of Plant Genomes*, edited by Bennetzen J., Flint-Garcia S., Hirsch C., Tuberosa R. Springer, Cham.

Sánchez González J. de J., J. A. Ruiz Corral, G. M. García, G. R. Ojeda, L. D. la C. Larios, *et al.*, 2018 Ecogeography of teosinte. PLoS One 13: e0192676. https://doi.org/10.1371/journal.pone.0192676

Schultz S. T., and J. H. Willis, 1995 Individual variation in inbreeding depression: The roles of inbreeding history and mutation. Genetics 141: 1209–1223. https://doi.org/10.1093/genetics/141.3.1209

Seabold S., and J. Perktold, 2010 statsmodels: Econometric and statistical modeling with Python, in *Proceedings of the 9th Python in Science Conference*.

Shaw R. G., D. L. Byers, and F. H. Shaw, 1998 Genetic components of variation in Nemophila menziesii undergoing inbreeding: Morphology and flowering time. Genetics 150: 1649–1661.

Simons Y. B., M. C. Turchin, J. K. Pritchard, and G. Sella, 2014 The deleterious mutation load is insensitive to recent population history. Nat. Genet. 46: 220–224. https://doi.org/10.1038/ng.2896

Swanson-Wagner R., R. Briskine, R. Schaefer, M. B. Hufford, J. Ross-Ibarra, *et al.*, 2012 Reshaping of the maize transcriptome by domestication. Proc. Natl. Acad. Sci. U. S. A. 109: 11878–11883. https://doi.org/10.1073/pnas.1201961109

Troth A., J. R. Puzey, R. S. Kim, J. H. Willis, and J. K. Kelly, 2018 Selective trade-offs maintain alleles underpinning complex trait variation in plants. Science 361: 475–478.

https://doi.org/10.1126/science.aat5760

Vaser R., S. Adusumalli, S. N. Leng, M. Sikic, and P. C. Ng, 2016 SIFT missense predictions for genomes. Nat. Protoc. 11: 1–9. https://doi.org/10.1038/nprot.2015.123

Waller D. M., J. Dole, and A. J. Bersch, 2008 Effects of stress and phenotypic variation on inbreeding depression in Brassica rapa. Evolution (N. Y). 62: 917–931. https://doi.org/10.1111/j.1558-5646.2008.00325.x

Waller D. M., 2021 Addressing Darwin' s dilemma : Can pseudo-overdominance explain persistent inbreeding depression and load? Evolution (N. Y). https://doi.org/10.1111/evo.14189

Walsh B., and M. Lynch, 2018 *Evolution and Selection of Quantitative Traits*. Oxford University Press, Oxford, UK.

Wang L., T. M. Beissinger, A. Lorant, C. Ross-Ibarra, J. Ross-Ibarra, *et al.*, 2017 The interplay of demography and selection during maize domestication and expansion. Genome Biol. 18: 215. https://doi.org/10.1186/s13059-017-1346-4

Wardyn B. M., J. W. Edwards, and K. R. Lamkey, 2007 The genetic structure of a maize population: The role of dominance. Crop Sci. 47: 467–474. https://doi.org/10.2135/cropsci2006.05.0294

Whitlock M. C., and K. Fowler, 1999 The changes in genetic and environmental variance with inbreeding in Drosophila melanogaster. Genetics 152: 345–353. https://doi.org/10.1093/genetics/152.1.345

Yang J., S. Mezmouk, A. Baumgarten, E. S. Buckler, K. E. Guill, *et al.*, 2017a Incomplete dominance of deleterious alleles contributes substantially to trait variation and heterosis in maize. PLoS Genet. 13: e1007019.

Yang N., X. W. Xu, R. R. Wang, W. L. Peng, L. Cai, *et al.*, 2017b Contributions of Zea mays subspecies mexicana haplotypes to modern maize. Nat. Commun. 8: 1–10. https://doi.org/10.1038/s41467-017-02063-5

Yang C. J., L. F. Samayoa, P. J. Bradbury, B. A. Olukolu, W. Xue, *et al.*, 2019 The genetic architecture of teosinte catalyzed and constrained maize domestication. Proc. Natl. Acad. Sci. U. S. A. 116: 5643–5652. https://doi.org/10.1073/pnas.1820997116

Zhang C., P. Wang, D. Tang, Z. Yang, F. Lu, *et al.*, 2019 The genetic basis of inbreeding depression in potato. Nat. Genet. 51: 374–378.

Figures

Figure 1. Percent inbreeding depression, regression coefficient of individual trait values on individual genomic inbreeding estimates, and ratio of dominance to total genetic variance for 18 traits measured in maize and teosinte.

Figure 2. Predicted segregating mutational burden per parent in the parents of maize and teosinte populations. (A) Total number of sites heterozygous or homozygous for segregating deleterious alleles per individual parent, (B) Mean burden per site per parent based on genomic evolutionary rate profiling (GERP) score under a model of partial recessivity, including only sites segregating within the parent's population, (C) Total burden per parent based on GERP scores, (D) Derived allele frequency spectrum for maize and teosinte populations. A total of 293,720 and 362,145 variants with GERP scores are segregating within maize and teosinte, respectively.

Figure 3. Variation in outbred and selfed progeny breeding values among parents in teosinte and maize. Mean values for self-fertilized S1 progenies (Y-axis) and outbred progenies (X-axis) scaled to the outbred population mean for each parent. Each trait is plotted in a different color. The 1:1 line is plotted in black, deviations from this line in the vertical direction correspond to the inbreeding depression for a single parent's selfed prognies compared to their outcrossed siblings.

Figure 4. Rare allele linkage scan for five traits in maize landrace and teosinte populations. Each column represents one chromosome; linkage map cM positions are displayed on the x-axis. Each row represents LOD scores of linkage scan for one trait. Each founder parent is represented by a different color and the two different haplotypes of each parent are distinguished by solid vs. dashed lines. Red solid line indicates the whole-genome LOD significance threshold.

Figure 5. The proportion of trait variance due to rare allele scan QTL (red color) and polygenic background effects (blue color) for each of 18 traits measured in maize and teosinte populations. The number of rare allele scan QTL detected for each population and trait is indicated on top of bars.

Figure 6. Segregation distortion whole genome scan results plotted as $-\log_{10} p$-value of the chi-square test for agreement to expected 1:2:1 segregation at each genome region (cM position) for (A) three selfed families in maize, and (B) eleven selfed families in teosinte; each family scan is plotted on a different row. Blue arrows indicate significant segregation distortion regions.

Figure 7. Percent inbreeding depression of individual families related to the parental rare allele load (RAL) measured genome-wide and within QTL CI for affected traits in maize and teosinte. Only traits for which a significant ($P < 0.05$) correlation was observed within one of the populations are displayed.

Figure 1. Percent inbreeding depression, regression coefficient of individual trait values on individual genomic inbreeding estimates, and ratio of dominance to total genetic variance for 18 traits measured in maize and teosinte.

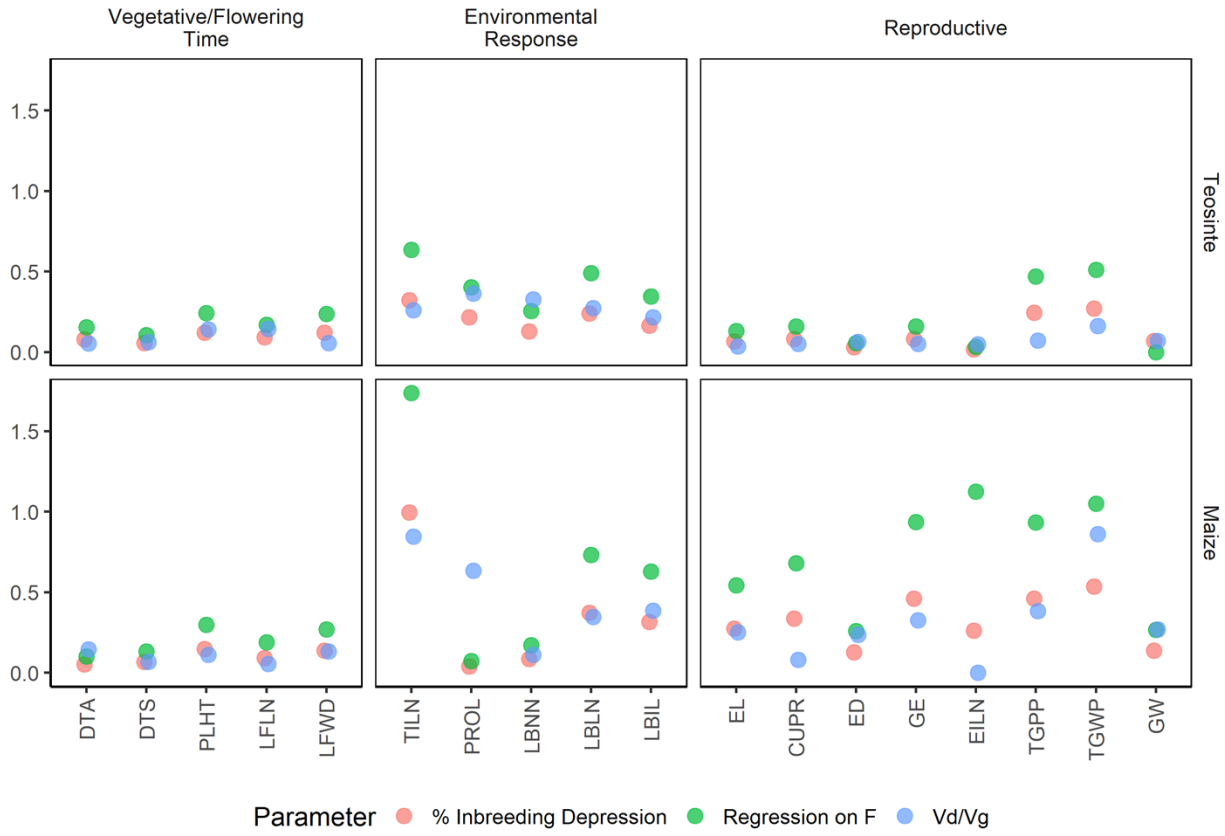


Figure 2. Predicted segregating mutational burden per parent in the parents of maize and teosinte populations. (A) Total number of sites heterozygous or homozygous for segregating deleterious alleles per individual parent, (B) Mean burden per site per parent based on genomic evolutionary rate profiling (GERP) score under a model of partial recessivity, including only sites segregating within the parent's population, (C) Total burden per parent based on GERP scores, (D) Derived allele frequency spectrum for maize and teosinte populations. A total of 293,720 and 362,145 variants with GERP scores are segregating within maize and teosinte, respectively.

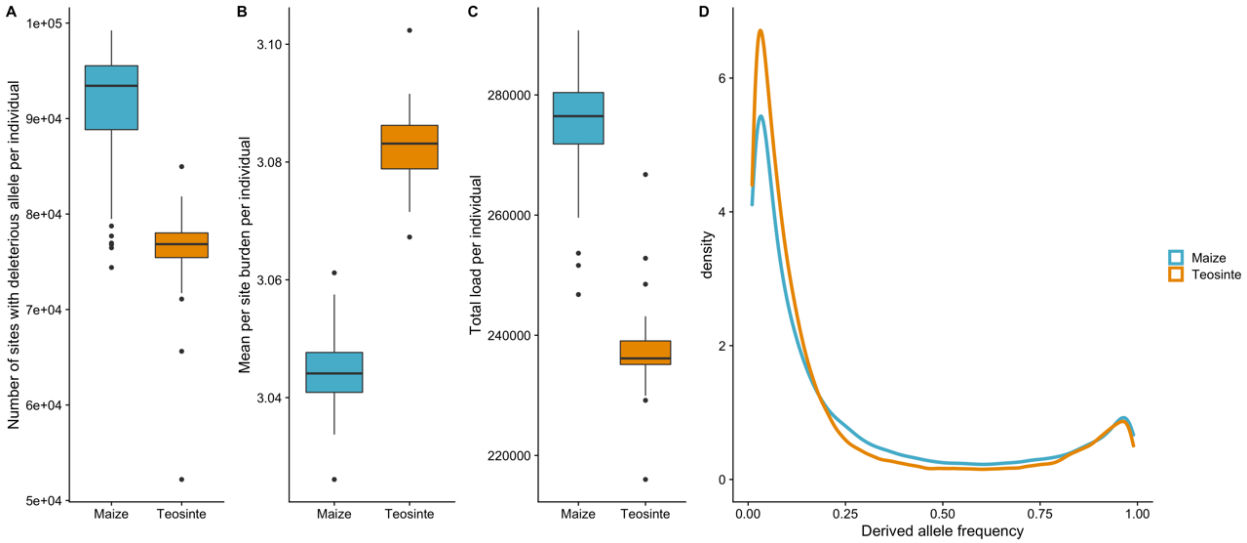


Figure 3. Variation in outbred and selfed progeny breeding values among parents in teosinte and maize. Mean values for self-fertilized S1 progenies (Y-axis) and outbred progenies (X-axis) scaled to the outbred population mean for each parent. Each trait is plotted in a different color. The 1:1 line is plotted in black, deviations from this line in the vertical direction correspond to the inbreeding depression for a single parent's selfed progenies compared to their outcrossed siblings.

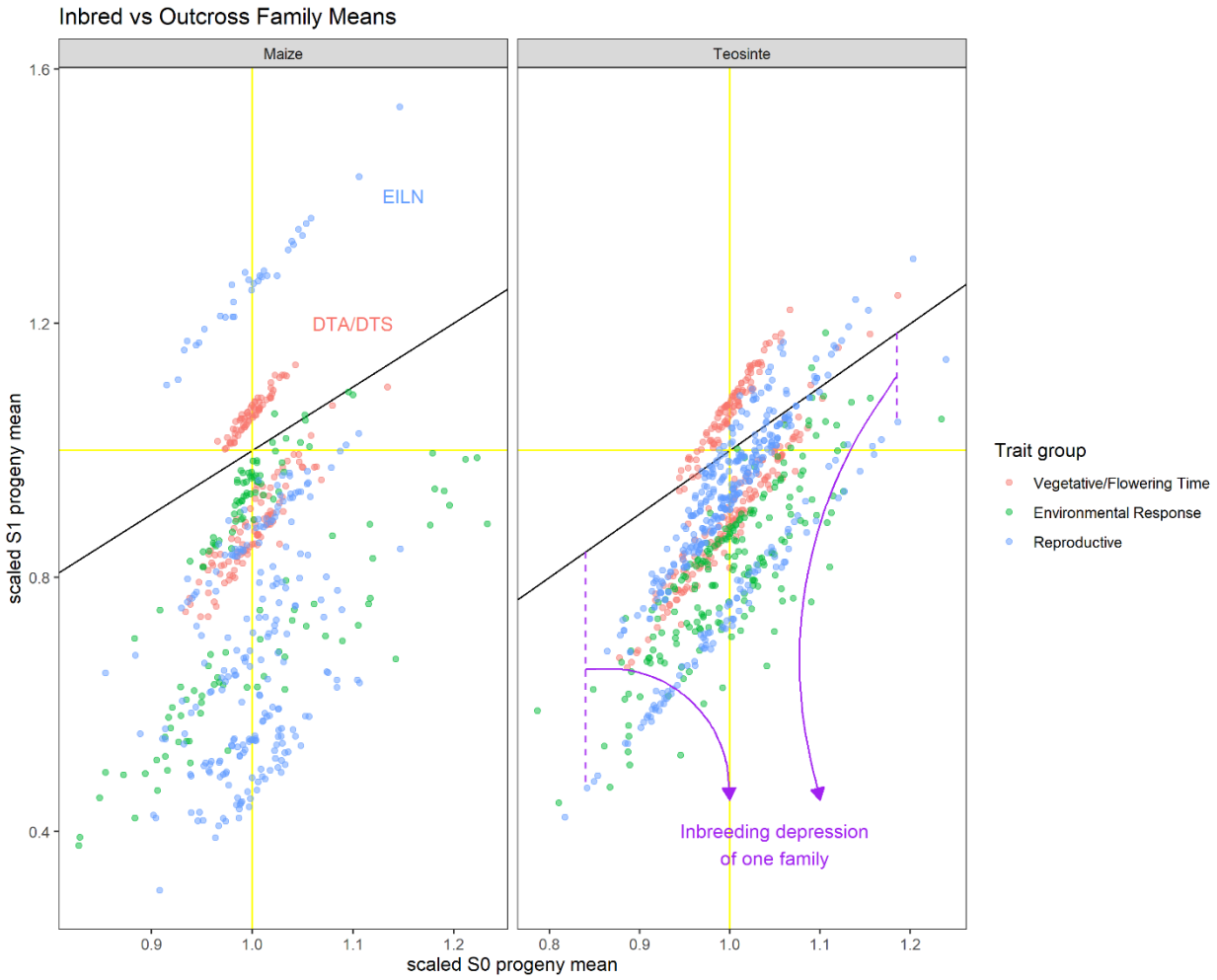


Figure 4. Rare allele linkage scan for five traits in maize landrace and teosinte populations. Each column represents one chromosome; linkage map cM positions are displayed on the x-axis. Each row represents LOD scores of linkage scan for one trait. Each founder parent is represented by a different color and the two different haplotypes of each parent are distinguished by solid vs. dashed lines. Red solid line indicates the whole-genome LOD significance threshold.

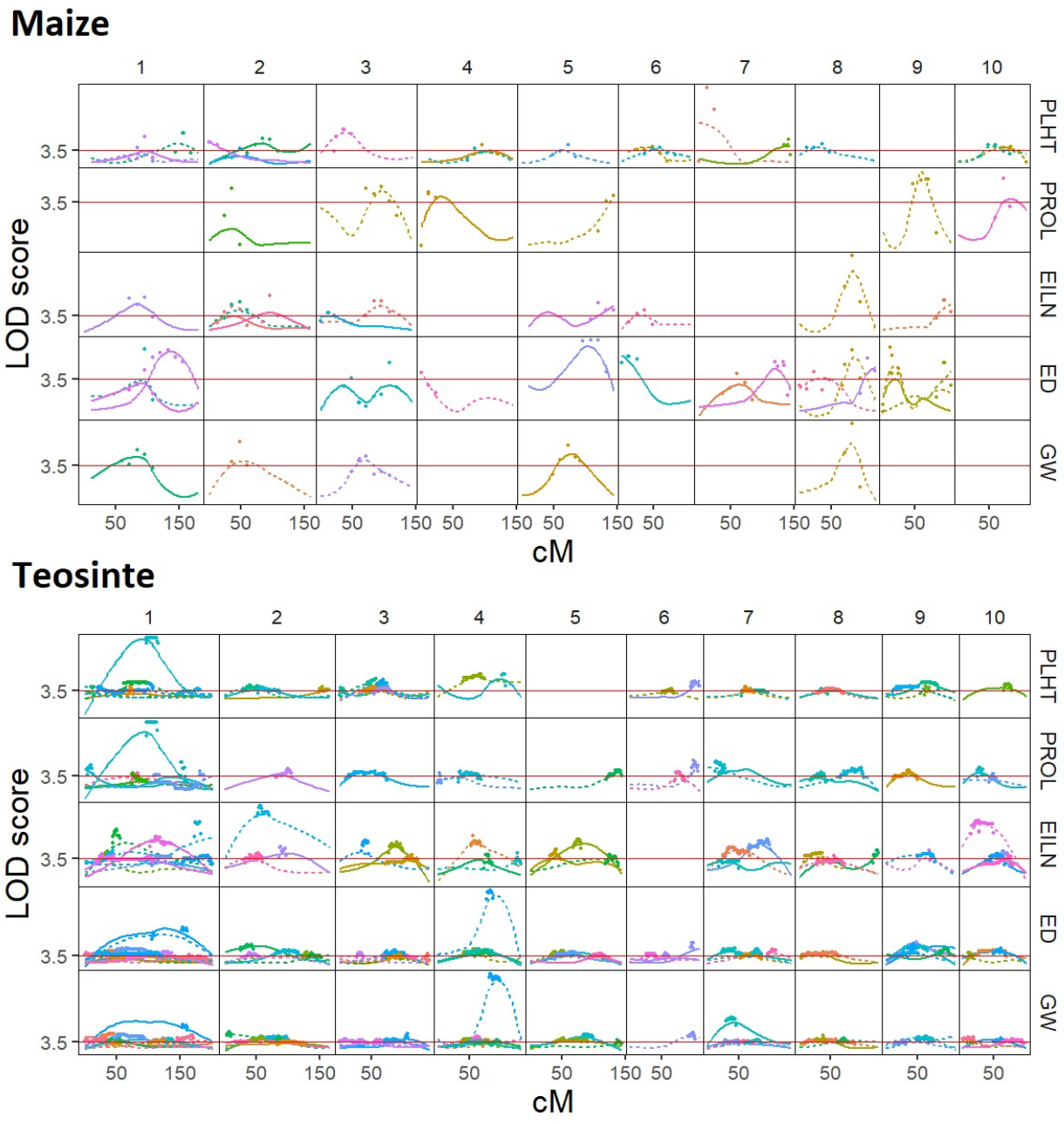


Figure 5. The proportion of trait variance due to rare allele scan QTL (red color) and polygenic background effects (blue color) for each of 18 traits measured in maize and teosinte populations. The number of rare allele scan QTL detected for each population and trait is indicated on top of bars.

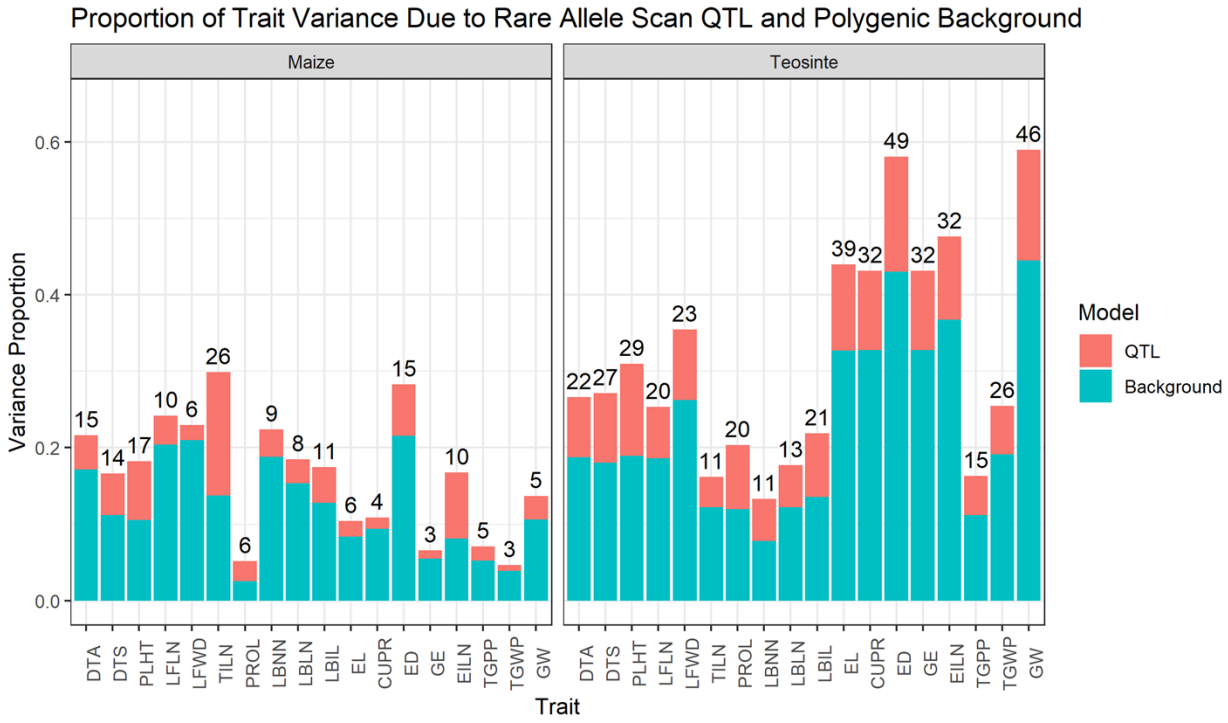


Figure 6. Segregation distortion whole genome scan results plotted as -$\log_{10}$ *p*-value of the chi-square test for agreement to expected 1:2:1 segregation at each genome region (cM position) for (A) three selfed families in maize, and (B) eleven selfed families in teosinte; each family scan is plotted on a different row. Blue arrows indicate significant segregation distortion regions.

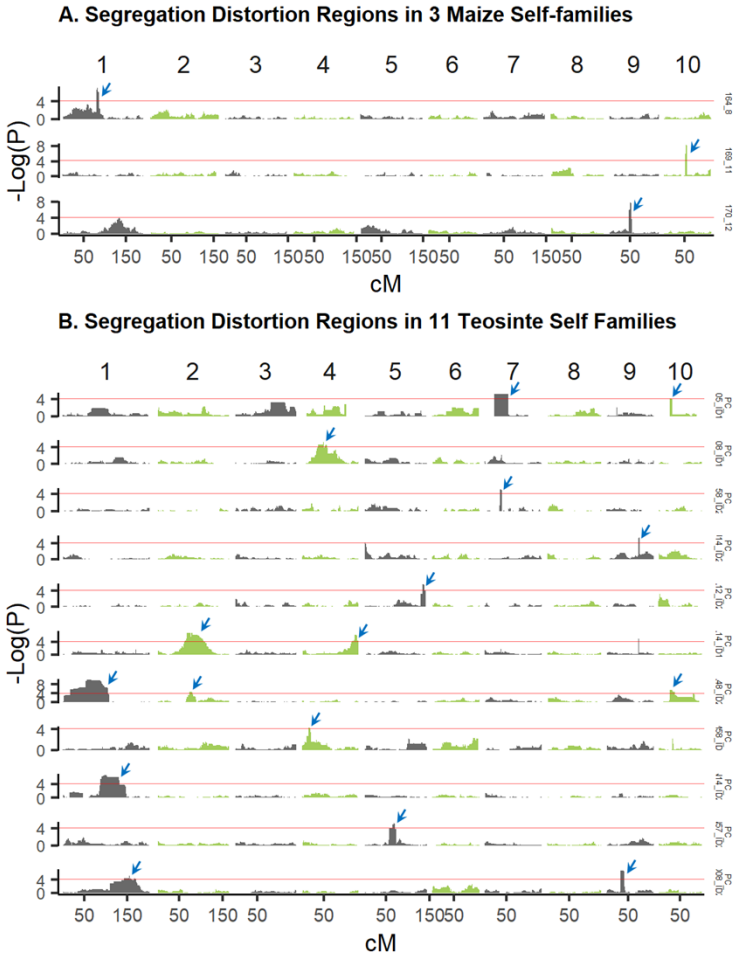
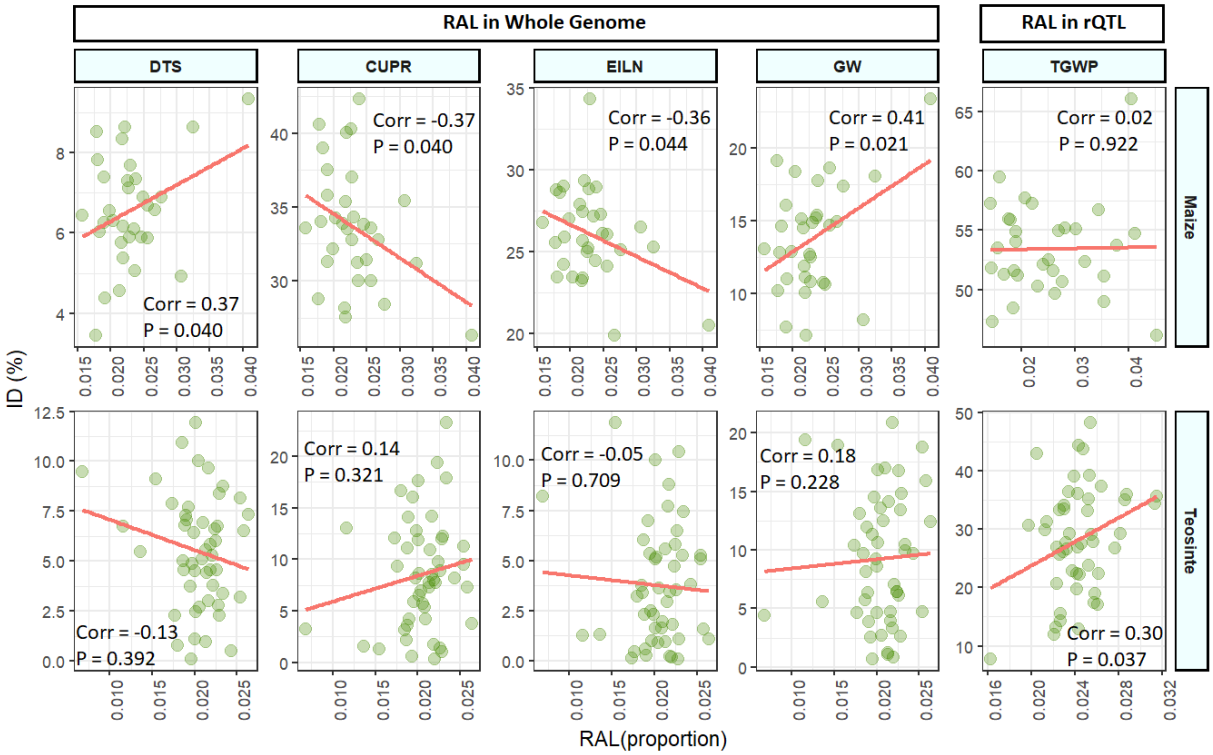


Figure 7. Percent inbreeding depression of individual families related to the parental rare allele load (RAL) measured genome-wide and within QTL CI for affected traits in maize and teosinte. Only traits for which a significant ($P < 0.05$) correlation was observed within one of the populations are displayed.



Tables

Table 1. Trait Abbreviations
List of 18 teosinte-maize landrace comparable traits and the corresponding acronyms, units and trait groups. From Yang et al. (2019).

Trait	Acronym	Units	Trait Group
Days to Anthesis	DTA	days	Veg/FT
Days to Silking	DTS	days	Veg/FT
Plant Height	PLHT	cm	Veg/FT
Leaf Length	LFLN	cm	Veg/FT
Leaf Width	LFWD	cm	Veg/FT
Tiller Number	TILN	count	EnvRes
Prolificacy	PROL	count	EnvRes
Lateral Branch Node Number	LBNN	count	EnvRes
Lateral Branch Length	LBLN	mm	EnvRes
Lateral Branch Internode Length	LBIL	mm	EnvRes
Ear Length	EL	mm	Rep
Cupules per Row	CUPR	count	Rep
Ear Diameter	ED	mm	Rep
Grains Per Ear	GE	count	Rep
Ear Internode Length	EILN	mm	Rep
Total Grain per Plant	TGPP	count	Rep
Total Grain Weight per Plant	TGWP	g	Rep
Grain Weight	GW	mg	Rep

Supplemental Figures

Figure S1. Diagram of key features of rare allele linkage scan compared to genome wide association study tests for quantitative trait loci.

Figure S2. Mean value of self-fertilized progenies as a proportion of outbred progeny means for 18 traits in teosinte and maize.

Figure S3. Predicted segregating mutational burden per parent in the parents of maize and teosinte populations. (A) Total segregating deleterious sites per individual parent, (B) homozygous deleterious sites per parent, (C) Heterozygous deleterious sites per parent, (D) Mean burden per site per parent based on genomic evolutionary rate profiling (GERP) score under a model of partial recessivity, including only sites segregating within the parent's population, (E) Mean homozygous burden per parent, (F) Mean heterozygous burden per parent, (G) Total burden per parent based on GERP scores, (H) Total homozygous burden per parent based on GERP scores, (I) Total heterozygous burden per parent.

Figure S4. Distribution of genomic inbreeding coefficient estimates among individuals in maize and teosinte populations.

Figure S5. Ratio of observed variance among S1 families to the expected variance based on $\sigma_A^2 + 0.25\sigma_D^2$, using estimates of σ_A^2 and σ_D^2 from outbred progenies and ignoring additional non-additive components of covariance introduced under inbreeding. Values greater than one are evidence for the importance of epistatic or D_2 variances, or a positive D_1 covariance. In contrast, if the ratio of the observed to predicted S_1 family variance is less than one, it is evidence for a negative D_1 covariance.

Figure S6. Additive (Va) and dominance (Vd) variances and covariance between additive and homozygous dominance deviations (D1) for one locus with favorable allele gene action effect a = 1, dominance gene action of d = 0, 0.5, 1, or 2, and varying allele frequency (p).

Figure S7. Proportion (R^2) of variation associated with genetic background (parentage estimated by principal components of the genome-wide marker data) and rare allele scan QTL effects for vegetative and flowering traits in maize and teosinte. Full model includes parentage, QTL additive, and QTL dominance effects. The proportion of variance due specifically to parentage, additive plus dominance QTL effects, additive QTL effects only, or dominance QTL effects only was estimated by measuring the decrease in R^2 after removing one of those factors from the full model.

Figure S8. Proportion (R^2) of variation associated with genetic background (parentage estimated by principal components of the genome-wide marker data) and rare allele scan QTL effects for environmental response traits in maize and teosinte. Full model includes parentage, QTL additive, and QTL dominance effects. The proportion of variance due specifically to parentage, additive plus dominance QTL effects, additive QTL effects only, or dominance QTL effects only was estimated by measuring the decrease in R^2 after removing one of those factors from the full model.

Figure S9. Proportion (R^2) of variation associated with genetic background (parentage estimated by principal components of the genome-wide marker data) and rare allele scan QTL effects for reproductive

traits in maize and teosinte. Full model includes parentage, QTL additive, and QTL dominance effects. The proportion of variance due specifically to parentage, additive plus dominance QTL effects, additive QTL effects only, or dominance QTL effects only was estimated by measuring the decrease in R^2 after removing one of those factors from the full model.

Figure S10. Relationships between the number of rare allele scan QTL detected per parent and the progeny sample size of each parent for teosinte and maize.

Figure S11. Proportion of traits with significant clustering of QTL between maize and teosinte populations, between QTL and significant GWAS associations, and between QTL detected in different parents within populations.

Figure S12. Rare allele scan results for vegetative and flowering traits in maize. Each row of figures corresponds to one trait. Each column corresponds to one of the ten chromosome pairs in maize. Logarithm of odds (LOD) scores for QTL models are plotted for the 2-LOD support interval for each QTL. LOD curves correspond to the effects of a single parental haplotype, and different parent effects are plotted with different colors. Blue dots represent the -log10 *p*-values of significant GWAS associations from Chen et al. (2020).

Figure S13. Rare allele scan results for environmental response traits in maize. Each row of figures corresponds to one trait. Each column corresponds to one of the ten chromosome pairs in maize. Logarithm of odds (LOD) scores for QTL models are plotted for the 2-LOD support interval for each QTL. LOD curves correspond to the effects of a single parental haplotype, and different parent effects are plotted with different colors. Blue dots represent the -log10 *p*-values of significant GWAS associations from Chen et al. (2020).

Figure S14. Rare allele scan results for reproductive traits in maize. Each row of figures corresponds to one trait. Each column corresponds to one of the ten chromosome pairs in maize. Logarithm of odds (LOD) scores for QTL models are plotted for the 2-LOD support interval for each QTL. LOD curves correspond to the effects of a single parental haplotype, and different parent effects are plotted with different colors. Blue dots represent the -log10 *p*-values of significant GWAS associations from Chen et al. (2020).

Figure S15. Rare allele scan results for vegetative and flowering traits in teosinte. Each row of figures corresponds to one trait. Each column corresponds to one of the ten chromosome pairs in maize/teosinte. Logarithm of odds (LOD) scores for QTL models are plotted for the 2-LOD support interval for each QTL. LOD curves correspond to the effects of a single parental haplotype, and different parent effects are plotted with different colors. Blue dots represent the -log10 *p*-values of significant GWAS associations from Chen et al. (2020).

Figure S16. Rare allele scan results for environmental response traits in teosinte. Each row of figures corresponds to one trait. Each column corresponds to one of the ten chromosome pairs in maize/teosinte. Logarithm of odds (LOD) scores for QTL models are plotted for the 2-LOD support interval for each QTL. LOD curves correspond to the effects of a single parental haplotype, and different

parent effects are plotted with different colors. Blue dots represent the -log10 *p*-values of significant GWAS associations from Chen et al. (2020).

Figure S17. Rare allele scan results for reproductive traits in teosinte. Each row of figures corresponds to one trait. Each column corresponds to one of the ten chromosome pairs in maize/teosinte. Logarithm of odds (LOD) scores for QTL models are plotted for the 2-LOD support interval for each QTL. LOD curves correspond to the effects of a single parental haplotype, and different parent effects are plotted with different colors. Blue dots represent the -log10 *p*-values of significant GWAS associations from Chen et al. (2020).

Figure S18. Relationship between additive (red) and dominance (blue) effect estimates associated with rare allele scan (RAS) QTL (X-axis) and corresponding GWAS associations inside the QTL intervals (Y-axis) in maize and teosinte. Effects are standardized to the phenotypic standard deviation for each trait and population.

Figure S19. Relationship between standardized additive and dominance effect estimates of rare allele scan QTL. Effects are standardized to the phenotypic standard deviation for each trait and population.

Figure S20. Standardized additive (a) and dominance (d) effects of rare allele scan QTL effects plotted by genome position within each of the ten chromosome pairs for both maize (A) and teosinte (B). Effects are standardized to the phenotypic standard deviation for each trait and population. Effects are plotted within trait categories. Right hand panels ("Eff_Direct") show the proportion of QTL allele effects where the rare variant effect is favorable (against the direction of inbreeding depression) or unfavorable (in the same direction as inbreeding depression) within categories based on the level of dominance of the rare allele, summed over all traits. Favorable allele effects are (partially to fully) recessive when a > 0 and d < 0, (partially to fully) dominant when a > 0 and d < 0, and overdominant when d > a > 0. Unfavorable allele are (partially to fully) recessive when a < 0 and d > 0, (partially to fully) dominant when a < 0 and d < 0 alleles, and overdominant when d < a < 0.

Figure S21. Proportion of trait inbreeding depression predicted based on rare allele scan (RAS) or genome-wide association study (GWAS) QTL dominance effect estimates and allele frequencies ($\sum_i^N p_i(1 - p_i)d_i$) for each trait in maize and teosinte.

Figure S22. Observed selfed progeny genotype frequencies at loci detected as segregation distortion regions in maize and teosinte. Genotypes *aa* and *bb* refer to homozygotes for one of the parental alleles, *ab* refers to heterozygotes. Shapes refer to particular families.

Figure S23. Segregation distortion regions (SDR) superimposed on physical linkage maps of maize (A) or teosinte (B). The local recombination rate (cM/Mb) is plotted as intensity of red color for each 1-cM window within either population. The positions of *gametophyte factor 1* (*Ga1*), *gametophyte factor 2* (*Ga2*), and *teosinte crossing barrier 1* (*Tcb1*) loci are indicated in purple

Figure S24. Segregation distortion regions (SDR; blue bars) superimposed on physical linkage maps of teosinte chromosomes also carrying seed dormancy QTL (green diamonds) detected by Chen et al.

(2020). The local recombination rate (cM/Mb) is plotted as intensity of red color for each 1-cM window within either population. The position of *gametophyte factor 2* (*Ga2*) locus is indicated in purple (this locus has not been finely mapped to date).

Figure S25. Observed selfed progeny genotype frequencies at loci detected as segregation distortion regions and overlapping known gametophyte factors in teosinte. Genotypes *aa* and *bb* refer to homozygotes for one of the parental alleles, *ab* refers to heterozygotes. Shapes refer to particular families. Right hand bar indicates the position of SDR (blue bars) and gametophyte factors (purple) on chromosomes 4 and 5, with the specific parents giving rise to the SDR indicated within their SDR.

Figure S26. Minor allele frequency spectrum for SNPs derived from whole genome sequencing of maize (blue color bars) and teosinte (red color bars) parents.

Supplemental Tables

Table S1. Teosinte mating design details. The number of progenies from each maternal parent are in rows, the number of progenies from each paternal parent are in columns. Values on the diagonal correspond to the number of selfed progenies per parent.

Table S2. Maize mating design details. The number of progenies from each maternal parent are in rows, the number of progenies from each paternal parent are in columns. Values on the diagonal correspond to the number of selfed progenies per parent.

Table S3. Inbreeding depression summary results from diallel analysis presented in this paper and realized relationship matrix analyses presented by Yang et al (2019) for both teosinte and maize. Overall outbred progeny mean values (S0_mean), selfed progeny mean values (S1_mean), scaled S1 progeny means as a proportion of the S0 mean value (S1_mean_scaled), percent inbreeding depression ((S1_mean – S0_mean)/S0_mean; ID_perc), absolute value of percent inbreeding depression (ID_perc_abs), the slope of the regression of individual trait values on estimated genomic inbreeding coefficient (Beta), the absolute value of the regression slope as a proportion of outbred mean value (Beta_scaled), additive genetic variance estimated from diallel (Additive), additive genetic variance estimated from the realized relationship matrices (VA_realized), dominance genetic variance estimated from the diallel (Dominance), dominance genetic variance estimated from the realized relationship matrices (VD_realized), narrow-sense heritability estimated from the diallel (h2), narrow-sense heritability estimated from the realized relationship matrices (h2_realized), ratio of dominance to total genetic variation estimated from the diallel (D_G_ratio), ratio of dominance genetic variance to total phenotypic variance estimated from the diallel (D_P_ratio), ratio of dominance to total genetic variation estimated from the realized relationship matrices (D_G_ratio_realized), ratio of dominance genetic variance to total phenotypic variance estimated from the realized relationship matrices (D_P_ratio_realized), heritability of S1 family means (H_S1), error variance estimated from outbred progenies (ErrorS0), error variance estimated from selfed progenies (ErrorS1), ratio of S1 to S0 error variances (S1_S0_Resid), total phenotypic variance among outbred individuals (Pheno_out), total phenotypic variance among selfed individuals (Pheno_S1), genotypic variance among S1 families (S1_Var), ratio of observed S1 family genetic variance to expected S1 family variance estimated as $\sigma_A^2 + 0.25\sigma_D^2$, based on additive and dominance variances estimated from outbred progenies, correlation between outbred and selfed breeding values for each parent (rS0S1), correlation between outbred breeding value and inbreeding depression for each parent (rS0BV_ID), correlation between outbred breeding value and percent inbreeding depression for each parent (r_S0BV_IDperc).

Table S4. Genetic burden per parent by predicted deleterious genotypic class: parental code (individual), population, genotypic count at deleterious sites (anc_geno; 0 = homozygous for ancestral allele, 1 = heterozygous, 2 = homozygous for derived, putatively deleterious allele), number of sites within genotypic class (del_sites), mean burden score per site within genotypic class (mean_load), sum of load scores across sites within genotypic class (total_load).

Table S5. Correlations among teosinte population parameter estimates. Matrix of correlation coefficients among the parameters in Table S3 across traits.

Table S6. Correlations among maize population parameter estimates. Matrix of correlation coefficients among the parameters in Table S3 across traits.

Table S7. Outbred and inbred progeny mean values, and family-specific inbreeding depression for individual parents within maize and teosinte populations. S0_effect and S1_effect are family effects scaled to the overall outbred population mean.

Table S8. Rare allele scan QTL effects in teosinte and maize. SNP, parent name, and haplotype number for QTL peak joined together (Marker), proportion of trait variance due to additive effect of this haplotype vs. all others (r2_A), proportion of trait variance due to dominance effect of this haplotype (r2_D), proportion of trait variance due to the combined additive and dominance effects of this haplotype vs. all others (r2_AD), additive effect of this haplotype allele vs. all others (a_effect), p-value associated with the additive effect (a_pval), t-value associated with the additive effect (a_tval), dominance effect of this haplotype (d_effect), p-value associated with the dominance effect (d_pval), t-value associated with the dominance effect (d_tval), number of heterozygotes for this haplotype (N_het), number of homozygotes for this haplotype (N_hom), peak SNP, parent, haplotype number, chromosome, physical position of SNP on AGP version 4 of the B73 reference sequence (pos_Agpv4), cM position of SNP, SNP defining left side of QTL interval (Pos_start), genetic position of SNP defining left side of QTL support interval (cM_start), physical position of SNP defining left side of QTL support interval (Pos_start), SNP defining right side of QTL interval (SNP_end), genetic position of SNP defining right side of QTL support interval (cM_end), physical position of SNP defining right side of QTL support interval (Pos_end), population, trait group, and number of GWAS associations identified by Chen et al. (2020) for same trait and population inside the support interval (N.GWAS.hits).

Table S9. Pleiotropic rare allele scan QTL in maize. Filtered version of Table S8 including only overlapping haplotype QTL positions with effects on more than one trait in maize.

Table S10. Pleiotropic rare allele scan QTL in teosinte. Filtered version of Table S8 including only overlapping haplotype QTL positions with effects on more than one trait in teosinte.

Table S11. Overlapping rare allele scan QTL and GWAS associations. GWAS associations detected by Chen et al (2020) encompassed by rare allele scan QTL for the same trait and population. SNP, parent name, and haplotype number for rare allele scan QTL peak joined together (QTL), marker identified by GWAS (SNP), chromosome (chrom_QTL and chrom_GWAS), additive effect of GWAS association (a_effect_SNP), dominance effect of GWAS association (d_effect_SNP), AgpV4 physical position of GWAS associated SNP (Position), Std. Dev of trait within the population, and proportion of trait variance associated with GWAS SNP (r2_GWAS). Other trait columns are the same as for Table S8.

Table S12. Summary of overlapping QTL and GWAS associations. Comparisons include QTL between populations, which are overlapping QTL for the same trait mapped in both maize and teosinte; QTL within populations, which are overlapping QTL for the same trait and population, but identified as the

contrast between different specific parental alleles and the rest of the population alleles; GWAS hits within QTL, which are GWAS SNPs that are encompassed by a QTL for the same trait and population; and QTL containing GWAS SNPs, which are QTL that encompass a GWAS SNP for the same trait and population. Each comparison is quantified by the mean number of overlaps per comparison (Mean.overlaps), the number of comparisons with zero overlaps (N.zero), the empirical permutation-based *p*-value for mean overlaps (Means.overlaps.pval), the empirical permutation-based p-value for the number of comparisons with zero overlaps (N.zero.overlaps.pval), and the total number of QTL or GWAS associations that represent the 'denominator' of these comparisons (N.QTL.or.hits). In addition, a Boolean variable (Clustered) indicates whether the observed mean number of overlaps is greater than the mean number obtained from the permutation test; this is not a significance test, if the empirical *p*-value indicates that the observed number of overlaps is significantly different than the random distribution, the variable Clustered indicates evidence of clustering (True) or overdispersion (False).

Table S13. Summary of segregation distortion analysis in maize and teosinte, indicating the parent of the self-fertilized family where segregation distortion was detected (Family); the chromosome (Chr); the SNP name (marker_start), cM position (cM_start), AgpV4 position (pos_start(bp) and pos_start(Mb)) of the SNP defining the left side of the segregation distortion region; the SNP name (marker_end), cM position (cM_end), AgpV4 position (pos_end(bp) and pos_end(Mb)) of the SNP defining the right side of the segregation distortion region; the length of the region in bases (Length(Mb)) and in cM (Length(cM)), the recombination rate within the region (cM/Mbp; SDR-RR), and the average recombination rate in the chromosome (CRR).

Table S14. Correlations between rare allele load across the whole genome and inbreeding depression for 18 traits in maize and teosinte.

Table S15. Correlations and p-values between rare allele load within QTL regions and inbreeding depression for 18 traits in maize and teosinte.

Table S16. Distribution of linkage disequilibrium correlation (*r*) values among parental gametes for minor alleles at adjacent marker pairs common between maize and teosinte with less than 10% missing values and less than 5% minor allele frequency (or between 5-10% minor allele frequency).