The genetic architecture and genomic context of glyphosate resistance

Kreiner, J.M.[*,1], Tranel, P.J.[2], Weigel, D.[3], Stinchcombe, J.R.[1,4], Wright, S.I.[1,5]

[1]*Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON M5S 3B2, Canada;*
[2]*Department of Crop Sciences, University of Illinois at Urbana–Champaign, Urbana, IL 61801*
[3]*Department of Molecular Biology, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany*
[4]*Koffler Scientific Reserve, University of Toronto, King City, ON L7B 1K5, Canada*
[5]*Centre for the Analysis of Genome Evolution and Function, University of Toronto, Toronto, ON M5S 3B2, Canada*

**Corresponding Author:* julia.kreiner@mail.utoronto.ca

Abstract

Although much of what we know about the genetic basis of herbicide resistance has come from detailed investigations of monogenic adaptation at known target-sites, the importance of polygenic resistance has been increasingly recognized. Despite this, little work has been done to characterize the genomic basis of herbicide resistance, including the number and distribution of involved genes, their effect sizes, allele frequencies, and signatures of selection. Here we implement genome-wide association (GWA) and population genomic approaches to examine the genetic architecture of glyphosate resistance in the problematic agricultural weed, *Amaranthus tuberculatus*. GWA correctly identifies the gene targeted by glyphosate, and additionally finds more than 100 genes across all 16 chromosomes associated with resistance. The encoded proteins have relevant non-target-site resistance and stress-related functions, with potential for pleiotropic roles in resistance to other herbicides and diverse life history traits. Resistance-related alleles are enriched for large effects and intermediate frequencies, implying that strong selection has shaped the genetic architecture of resistance despite potential pleiotropic costs. The range of common and rare allele involvement implies a partially shared genetic basis of non-target-site resistance across populations, complemented by population-specific alleles. Resistance-related alleles show evidence of balancing selection, and suggest a long-term maintenance of standing variation at stress-response loci that have implications for plant performance under herbicide pressure. By our estimates, genome-wide SNPs explain a comparable amount of the total variation in glyphosate resistance to monogenic mechanisms, indicating the potential for an underappreciated polygenic contribution to the evolution of herbicide resistance in weed populations.

Introduction

In theory, it should be easy to understand the genetic basis of herbicide resistance in weeds, because herbicides typically intentionally target a known biochemical pathway, and specific genes or gene products. In practice, however, herbicide resistance in evolved weed populations is complex. Agricultural weed populations evolve cross-resistance to multiple “modes” of

herbicides (distinct classes of herbicides designed to target select pathways) not only through well characterized, large-effect mutations that alter the interaction of the herbicide with the target enzyme (target-site resistance, TSR), but also through other mutations across the genome (non-target-site resistance, NTSR), often with smaller individual effects (Délye, 2013; Kreiner, Stinchcombe, & Wright, 2018; Tranel, Riggins, Bell, & Hager, 2011). While much work has been done to elucidate causative TSR mutations and their frequency in experimental and natural weed populations (Heap, 2010), we know little about the number of loci involved and the associated distribution of allelic effect sizes, frequencies, and the fraction of phenotypic variance explained by NTSR herbicide resistance loci. Here we combine genome-wide association approaches with population genomics to dissect the genetic architecture and genomic context of glyphosate resistance.

Efforts to better characterize the genetic basis of NTSR will help discover genetic markers for managing the spread of resistance. With a catalogue of genome-wide allelic effects on resistance, GWA methods can inform prediction of how populations, based on their genetics, may respond to selection from future herbicide pressures. More broadly, understanding the genetics of NTSR will contribute to our understanding of the relative importance and consistency of small- and large-effect alleles contributing to rapid adaptation, while also shedding light on how strong selection shapes genome-wide variation of natural weed populations.

It is likely that NTSR is as prevalent as TSR in conferring herbicide resistance across agriculturally important ranges (e.g. (Delye, Gardin, Boucansaud, Chauvel, & Petit, 2011)). Several studies have looked for causative mutations in the gene encoding the herbicide-targeted protein, but failed to find one—implying a widespread role of small effect mutations conferring resistance in these populations (Délye, 2013; J. Guo et al., 2015; Van Etten, Lee, Chang, & Baucom, 2019; Van Horn et al., 2018). In contrast to a single large-effect TSR allele, which is often likely to have arisen *de novo*, the polygenic basis of resistance with many small-effect alleles is likely to draw from genome-wide standing genetic variation (Kreiner et al., 2018; Neve, Vila-Aiub, & Roux, 2009). Thus NTSR mechanisms may even allow for naive populations to have some level of standing variation for resistance to herbicides. NTSR is a challenge for management not only because of its mysterious genetic basis and possible presence in naive

populations, but also due to pleiotropic effects of resistance alleles conferring cross-resistance to multiple herbicide modes (Preston, 2004; Preston, Tardif, Christopher, & Powles, 1996; Yu, Abdallah, Han, Owen, & Powles, 2009; Yu, Cairns, & Powles, 2007). With increased incidence of cross-resistance within individuals and populations (Heap, 2010), combined with reactive management through applications of tank mixes of different classes of herbicides, NTSR is likely to only increase in importance and prevalence.

A handful of gene families—encoding enzymes such as cytochrome P450s monooxygenases, glycosyltransferases, and glutathione S-transferases, or ABC transporters—have been repeatedly implicated in conferring NTSR, especially through the action of differential gene expression (Yuan, Tranel, & Stewart, 2007). These NTSR genes act by altering herbicide penetration, translocation, accumulation at the target site, as well as through offering protection from herbicide effects, the expression of which can be constitutive or induced (Délye, 2013; Moretti et al., 2018). Many examinations of potential NTSR genes have employed differential gene expression approaches, and several studies have successfully identified putative, diverse NTSR genes in such a manner (e.g. (Busi, Porri, Gaines, & Powles, 2018; I. Cummins, Cole, & Edwards, 1999; Duhoux, Carrère, Duhoux, & Délye, 2017; Duhoux, Carrère, Gouzy, Bonin, & Délye, 2015; Küpper, Peter, et al., 2018; W. Liu et al., 2018; Nakka et al., 2017; Pan et al., 2019; Peng et al., 2010; Varanasi, Brabham, & Norsworthy, 2018; N. Zhao et al., 2019)). Nonetheless, the genomic, rather than transcriptomic, basis of NTSR is unknown in most species and for most classes of herbicides (but see ((Ian Cummins et al., 2013; Van Etten et al., 2019).

While increasingly the genetic repeatability of herbicide resistance is being investigated with genomic approaches (Flood et al., 2016; Kreiner et al., 2019; Küpper, Manmathan, et al., 2018; Leslie & Baucom, 2014; Molin, Wright, Lawton-Rauh, & Saski, 2017; Van Etten et al., 2019), traditional genome-wide association (GWA) approaches have yet to be applied to explicitly characterize the genetic architecture of herbicide resistance, and the repeatability of alleles implicated across populations (but for related approaches, see (Benevenuto et al., 2019; Van Etten et al., 2019)). In other systems, GWAS have been widely and successfully used to identify new candidate genes, characterize the architecture of key traits, make predictions about phenotypes in studied and novel populations, and inform breeding strategies through genomic

selection (e.g. (Bock, Kantar, Caseys, Matthey-Doret, & Rieseberg, 2018; Epstein et al., 2018; Exposito-Alonso et al., 2019, 2018; Spindel et al., 2015; Swarts et al., 2017)).

Glyphosate based herbicides, commonly referred to by the brand name Round-up™, are widely used across North America to suppress weed populations, especially in soy and corn fields. Glyphosate, which targets a step in the synthesis of aromatic amino acids, carried out by 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), was commercialized in the 1970s with widespread deployment of glyphosate-resistant crops nearly 20 years later (W. G. Johnson, Davis, Kruger, & Weller, 2009). Since then, the incidence of glyphosate resistance (combined with multiple-resistance to other modes of herbicides) has increased exponentially, and is third to only ALS-inhibiting and Photosystem II-inhibiting herbicides in the number of resistant species (Heap, 2010). Meanwhile, glyphosate herbicides are still the most widely used across North America—hinged on low operational costs and glyphosate-resistant cropping systems—despite ubiquitous resistance in weed populations, especially in the genus *Amaranthus* (Duke & Powles, 2008; Vencill, Grey, Culpepper, Gaines, & Westra, 2008).

Amaranthus tuberculatus has been investigated by weed scientists for decades due to its problematic nature in agricultural fields and increasing reports of multiple-resistance (Bell, Hager, & Tranel, 2013; Bernards, Crespo, Kruger, Gaussoin, & Tranel, 2012; Matthew J. Foes, Tranel, Loyd M. Wax, & Edward W. Stoller, 1998; Patzoldt & Tranel, 2007; Patzoldt, Tranel, & Hager, 2002, 2005; Shergill, Bish, Jugulam, & Bradley, 2018; Tranel et al., 2011). More recently, a variety of genomic resources (a high quality reference genome, large resequencing dataset, phenotyping for glyphosate resistance) have been assembled (Kreiner et al., 2019), building on earlier efforts (Lee et al., 2009; Riggins, Peng, Stewart, & Tranel, 2010)). Recently, we characterized population structure, demographic history, and signals of monogenic selection at EPSPS and its related amplification in response to glyphosate, across natural and agricultural populations (Kreiner et al., 2019). With reports of non-target site resistance to glyphosate in the species (Nandula, Ray, Ribeiro, Pan, & Reddy, 2013), we next sought to identify genome-wide signals of adaptation to herbicides.

In this study, we implemented GWA and population genomic approaches for characterizing the genetic architecture of glyphosate resistance in *Amaranthus tuberculatus*. We used linear mixed models to identify new putative NTSR genes in addition to accurately identifying known, key large effect genes (TSR loci), and assess their relative importance. Our GWA for resistance to glyphosate in *Amaranthus tuberculatus* shows the largest genome-wide peak on chromosome 5 near the known-target, EPSPS. However, several other alleles show genome-wide significance after correcting for multiple testing (FDR), population structure, and after comparing to the permuted null-expectation. Genes encompassing these SNPs have relevant NTSR-related functions, and several candidates have been previously implicated in response to stress and herbicides. Furthermore, the associated SNPs are enriched for common alleles, large effects, and signals of selection. Our results imply a genetic architecture of NTSR resistance that makes use of both common alleles shared across populations and of rare population-specific alleles, with evidence of long term balancing selection being responsible for their diversity across populations.

Materials & Methods

Amaranthus tuberculatus resequencing & phenotype data

Resequencing and phenotype data were obtained from a published study (Kreiner et al., 2019), which included a high-quality female reference genome for the species. Whole-genome Illumina sequencing data are available at European Nucleotide Archive (ENA) (project no. PRJEB31711) (Kreiner, Giacomini, Bemm, Waithaka, Regalado, Lanz, Hildebrandt, Sikkema, Tranel, Weigel, Stinchcombe, Wright, 2019b), and the reference genome and its annotation available at CoGe (reference ID = 54057) (Kreiner, Giacomini, Bemm, Waithaka, Regalado, Lanz, Hildebrandt, Sikkema, Tranel, Weigel, Stinchcombe, Wright, 2019a). There were 158 agricultural samples, collected from 8 fields with high *A. tuberculatus* densities across Missouri and Illinois in the Midwest United States, and from newly infested counties in Ontario, Canada, Walpole Island and Essex. Lastly, 10 samples from naturally occurring, non-agricultural Ontario populations were included, totaling 168 resequenced samples.

As described in (Kreiner et al., 2019), phenotyping for glyphosate resistance was performed from offspring of field-collected specimens, and the same individuals were used for resequencing. Three weeks after a glyphosate application of 1,260 g glyphosate (WeatherMax 4.5 L; Monsanto) per hectare, plants were rated for resistance on a scale of 1 to 5 based on a visual scale of injury. We represented resistance as a continuous trait in our GWAs (i.e. the 1-5 scale), as continuous heritable variation in resistance has been previously reported (Holliday & Putwain, 1980; Patzoldt et al., 2002).

Linear mixed GWA models & preprocessing & GO enrichment

Filtered VCFs were obtained from (Kreiner et al., 2019) for all analyses. Briefly, freebayes-called SNPs were filtered based on missing data (>80% present), repeat content, allelic bias (<0.25 and >0.75), read paired status, and mapping quality (> Q30). Six individuals were removed due to excess missing data, leaving 162 for baseline analyses. Further investigations into identity-by-descent in these 162 samples showed some individuals with particularly high levels of relatedness, some of which was unaccounted for even when including a relatedness matrix in genome-wide association tests. We thus removed another 7 individuals, resulting in a final sample size of 155. A table of individuals by population, region, and their associated resistance-related phenotypes are included in the supplementary (**Sup. Table 1**).

We implemented several linear mixed model genome-wide association tests, and iteratively included two covariates with a causal genetic basis of glyphosate resistance. The first was a TSR mutation of proline to serine at codon number 106 in *EPSPS* (P106S). The second was the magnitude of the copy number of *EPSPS* (as characterized in (Kreiner et al., 2019)), based on the ratio of local coverage within the *EPSPS* gene to the median of genome-wide coverage. We did so in GEMMA (Zhou & Stephens, 2012), using the lmm -4 option and after estimation a kinship/relatedness matrix (-gk), with covariates specified (-c option) when relevant. Model selection was done holistically, by interpreting MLE, inflation factor for false positives and false negatives in qq plots, percent variance explained (PVE) by SNPs, and comparison of unpruned and LD-pruned SNP sets. We employed both a Bonferroni correction and an FDR approach to significance testing in R through the function p.adjust (method="bonferroni", "fdr"), where we used a false discovery rate of $\alpha = 0.05$. To calculate the significance thresholds, we took the raw,

$-\log_{10}$(p-value) equivalent to a q-value (FDR corrected p-value) < 0.05. While we show both, we used the FDR cutoff for significance testing as Bonferroni correction can often be overly conservative in a GWAS setting due to the assumption of independence between tests (R. C. Johnson et al., 2010), especially true in our case as we chose to not implement LD thinning. We only present results from the unpruned dataset here, as we found that LD-pruned sets generally had an excess of false negatives. To evaluate how well our best model (+TSR) did in accounting for variance in resistance due to EPSPS copy number, we estimated the r^2 of a regression of EPSPS copy number on genotypes of the 34 SNPs that fall within the EPSPS amplification and that are significantly correlated with resistance.

One challenge of interpreting GWAS analyses with covariates is that the percent variance explained (PVE) is calculated after the effects of the covariates are removed, making it hard to quantitatively compare the magnitude of the covariate's effects with SNP effects. As a point of comparison, we estimated a multiple regression model of glyphosate resistance ratings on TSR + EPSPS copy number. The overall model r^2 gives an indication of how much variation in resistance is accounted for by these monogenic effects, with 1- r^2 representing (approximately) the residual variances. We then took the residuals of this regression, after adding the mean resistance value to each observation to retain the same scale, and used it as our focal trait in a GWA—this should allow us to infer the percent of variance explained in phenotypic resistance, independent of monogenic mechanisms. In particular, it should allow us to infer the relative importance of monogenic to polygenic mechanisms by taking ((1-monogenic r^2) x Residual PVE) as the total amount of variance in phenotypic resistance explained by polygenic mechanisms.

We evaluated enrichment of and the roles of our significant SNPs (after controlling for relatedness and multiple test correction, FDR at the level of $\alpha = 0.05$) in PANTHER GO-Slim molecular functions, biological processes, cellular components, and protein classes. To do so, we identified *Arabidopsis thaliana* orthologues in the annotated female *Amaranthus tuberculatus* reference genome using OrthoFinder (Emms & Kelly, 2015) (as in (Kreiner et al., 2019)). Genes with significant SNPs where no *A. thaliana* orthologs were identified through orthofinder were manually curated from the reference genome annotation file (which was curated with annotations from several species to identify known genes, (Kreiner et al., 2019)). Finally, the set of manually

curated and OrthoFinder *A. thaliana* orthologues were used to evaluate molecular function and test for enrichment of particular gene classes in Panther GO-Slim. Enrichment of significant SNPs for all Panther categories was analyzed using a Fisher's exact test and FDR correction.

GWA permutations

To generate a null distribution of expected significant hits, and as a point of comparison for downstream population genetic analyses, we randomized phenotype with respect to genotypes. For permuted models that included covariates, we randomized the phenotype of interest and covariates together with respect to SNP genotypes, such that the associations between the resistance and covariates were maintained. We generated 250 sets of randomized data, and ran a GEMMA GWA on each randomized set. To generate null distributions of the expected number of significant SNPs, we used the significance threshold (q value < 0.05) estimated from each of the equivalent, observed GWAS models. From these iterations, we then calculated the average number of significant SNPs per permutation to get the FDR. For illustrative purposes, and to compare to the observed distribution of our true GWAS, we plotted the distribution of various statistics for all significant SNPs across all permuted GWASs (i.e., in **Fig 3**). For significance testing, we resampled the total pool of significant SNPs across all permuted GWASs to the same number of observations we observed in our true GWAS (or when relevant, the non Scaffold-5 SNP set), 1000 times. Each time, we took the median (or mean) of various statistics, and then the 5 and 95% quantile to attain the 95% CI of the empirical null distribution. To test whether there was enrichment compared to the null expectation, we then compared these CIs to the median (or mean) of these same statistics for the observed GWAS.

Summary statistics

We estimated F_{st} on a per-site basis for our set of putative NTSR alleles and genome-wide loci using VCFtools. To estimate F_{st} among resistant and susceptible individuals, we reclassified resistance as a binary trait (with individuals $\geq$ 2 on a scale of 1 to 5 classified as resistant) and calculated a per-site Weir-Cockerham F_{st} (Weir & Cockerham, 1984) among groups. For permuted sets, we made sure to calculate F_{ST} among groups of randomly assigned resistant and susceptible individuals. For 0-fold and 4-fold diversity in 10 kb windows around focal alleles, we used VCFtools per-site diversity estimates on VCF files containing variant and invariant 0-fold

or 4-fold sites, inferred in (Kreiner et al., 2019). Statistics for genomic windows were estimated with custom dplyr functions.

We also calculated two haplotype-based statistics, extended haplotype homozygosity (EHH, (Sabeti et al., 2002) and the integrated haplotype score (iHS). Both EHH and iHS are/can be calculated with respect to the alternate (1) and reference (0) allele haplotypes in selscan (Szpiech & Hernandez, 2014), and so to make these statistics specific to resistant versus susceptible comparison, we swapped allele assignment (1 versus 0) only for SNPs that had a negative effect on resistance, such that all 1 alleles correspond to alleles with positive effects on resistance. Both EHH and IHS calculations require phased haplotypes and a genetic map, and therefore we called phased haplotypes in SHAPEIT (Delaneau, Zagury, & Marchini, 2013), and estimated the population-based LD map using LDhat (McVean & Auton, 2007) (as in (Kreiner et al., 2019)). iHS calculations build on EHH by taking the difference in the EHH-distance curve among haplotypes and standardizing for allele frequency in bins across the genome, where haplotypes are differentially defined based on where EHH becomes < 0.05. To encompass all significant SNPs, we lowered the minor allele threshold to 0.01, as opposed to the default 0.05. Finally, we ran the program norm, implemented in selscan, to obtain the standardized iHS, which takes into account the genome-wide expected iHS value given a certain allele frequency. All custom scripts are available at https://github.com/jkreinz/NTSR-GWAS-MolEcol2020.

Results

SNPs and Traits

Our final SNP set from 155 *A. tuberculatus* individuals included 10,279,913 SNPs across 16 chromosome-resolved scaffolds. Of these, 8,496,628 SNPs were used by GEMMA, after sites with > 5% missing data and an allele frequency <1% were removed. The final dataset encompassed individuals with a range of resistance levels, *EPSPS* copy number, and a modest representation of TSR individuals. Of all individuals, 81 were classified as glyphosate resistant (a rating of ≥2 on a scale of 1-5) and 74 as susceptible (note that the continuous rate scale was used as input for the GWAS). Ten out of 155 individuals had a TSR mutation in the *EPSPS* coding sequence, while variance in *EPSPS* copy number was prevalent: 79 individuals had a

copy number ≤1.5, and the copy number in 76 individuals exceeded 1.5 (medians of 1.13 and 5.95 copies, respectively) with one individual ranging up to 30.

Gemma Linear Mixed Models

We compared four GWAS mixed effect models, all of which included a random effect kinship matrix but differed in the presence of fixed effect covariates: no covariates, TSR in the *EPSPS* coding region, *EPSPS* copy number, and both TSR and copy number changes (**Table 1**). Including covariates in a GEMMA model removes the effect of the covariate on the response variable before estimating SNP effects, and effectively makes the PVE the percent of residual variance explained by genome-wide SNPs (Zhou, Carbonetto, & Stephens, 2013). Accordingly, whether covariates were included in the model or not markedly altered its outcome. While the maximum likelihood estimate improved with the addition of covariates (TSR+Copy Number > Copy Number > TSR > no covariates), so did the inflation factor (lambda), indicating the typical trade-off between under- and overfitting. The ranking of significant SNPs based on q-values corresponding to a FDR of 5% suggested that a model including both covariates yielded the most significant hits, however this also corresponded with an estimated FDR more than 5-times higher than the TSR model (**Table 1**, see permuted median FDR). In contrast, the "Copy number" model seems to be underpowered (excess of false negatives) and simultaneously inflated for false positives. Only 17 SNPs pass the FDR correction genome wide, with a lambda of 1.07 and a qq plot that shows early departure from the observed:expected 1:1 line at $-\log_{10}(p) = 2$ (**Fig. 1, Table 1)**. Models that include "Copy number" as a covariate have the lowest explained PVE and atypical qq plots in terms of an enrichment of observed/expected p values just at the outer tail, as opposed to the model TSR and without covariates.

The GWA models without covariates and with TSR are the best in terms of an enrichment of observed, high $-\log_{10}$(p-value) sites, a critical factor for the identification of candidate loci and phenotypic prediction (**Table 1)**. The TSR model has 196 significant SNPs after correction, compared to 91 SNPs in the model without covariates. Despite a slight increase relative to the covariate free model, the TSR model is only estimated to have a 2.5% FDR (5/196 significant SNPs being false discoveries) and thus remains conservative for both variant discovery and phenotypic prediction (**Table 1)**. We will therefore refer predominantly to the TSR model for

310 further investigation of candidate genes and genetic architecture of resistance. However, it is important to note that by not having EPSPS copy number as a covariate in our focal models we expect that our SNP effects will include SNPs associated with gene copy amplification at this locus.

315 **Table 1.** GWA results from four linear mixed models varying in their fixed effects, based on the same genome-wide SNP set & kinship matrix as a random effect.

Covariates	**-log10(*p*) corresponding to $q = 0.05$**	**# Sig. SNPs @ FDR**	**PVE (SE)**	**MLE**	**Lambda (95% CI)**	**Permuted FDR estimate**
None	6.4	91	0.388 (0.230)	-268	0.979 (0.972, 0.986)	0.011 (1/91)
TSR	6.0	196	0.492 (0.238)	-264	0.983 (0.976, 0.990)	0.025 (5/196)
Copy Number	7.7	17	0.282 (0.197)	-245	1.073 (1.065, 1.081)	0.059 (1/17)
TSR + Copy Number	5.6	573	0.374 (0.219)	-234	0.979 (0.971, 0.986)	0.131 (75/573)

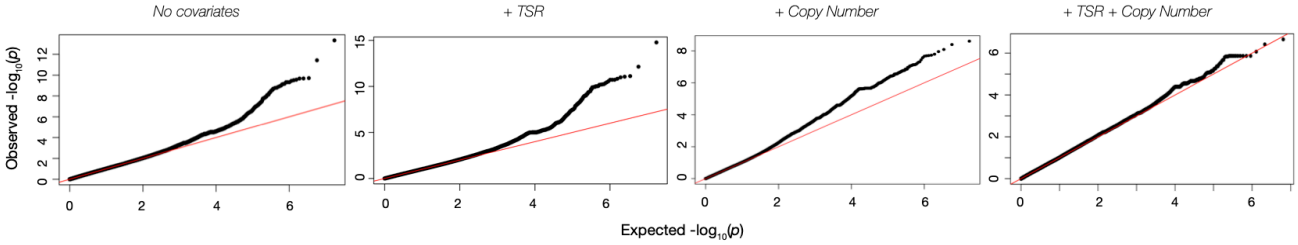


Fig 1. QQ plots of GWA results from four linear mixed models, varying in their fixed effects.

320 *Distribution of genome-wide associations & PVE*

The qualitative pattern of genome-wide associations with glyphosate resistance are similar among the covariate free and TSR model (**Fig. 2**). Zooming in on scaffold 5 of the genome assembly, in the center of which *EPSPS* is located (Kreiner et al., 2019), a butte of significant 325 SNPs appears within the confines of the EPSPS-related amplification (**Fig. 2b)**. These hits are not only localized to the 7 Mb extended region of *EPSPS* amplification —reassuringly, SNPs

significantly associated with glyphosate resistance are found directly within *EPSPS*. In both models, the *EPSPS* containing scaffold 5 shows the greatest number of significant SNPs across all scaffolds, almost four times more than any other in the TSR model (**Fig. 2c**). Thus, despite not including EPSPS copy number in our final model, variance in resistance related to this amplification appears to be accounted for. Indeed, the 34 significant resistance-associated SNPs that fall within the amplification can explain 76% of the variation in EPSPS copy number in a multiple regression framework ($F_{21,106} = 16.03$, $p < 2.2e\text{-}16$).

Although many significant SNPs fall within the *EPSPS* amplification (34/196) and two even within the *EPSPS* gene itself, our most significant SNP genome-wide is not located within *EPSPS* or the region of extended *EPSPS* amplification **(Fig 2a,b)**. The two most significant SNPs genome-wide are ~ 7 Mb downstream of *EPSPS* and ~2 Mb downstream of the region of extended *EPSPS* amplification, respectively (**Fig. 2b)**. It is in principle possible that this locus is linked to *EPSPS*, and better reflects expected signals of association at a diploid locus, compared to the high-copy number region within the *EPSPS*-related amplification. Alternatively, these SNPs may be linked to other loci with direct effects on glyphosate resistance, as these two SNPs tag *CYP76C6*, one of three cytochrome P450 genes significantly correlated with glyphosate resistance—a class of gene families consistently implicated in xenobiotic detoxification (Gaines et al., 2020; Yuan et al., 2007).

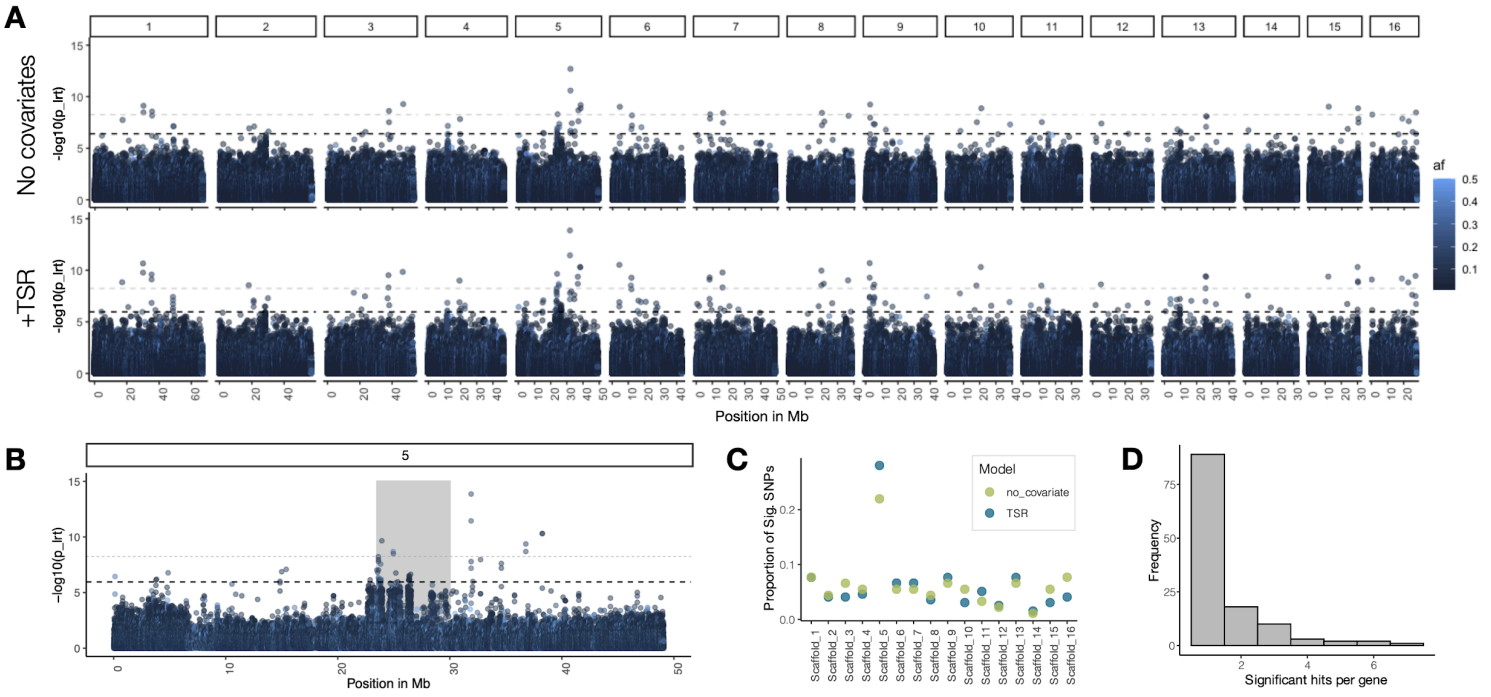


Fig. 2. GWA of glyphosate resistance. A) Manhattan plot of significance of association with glyphosate resistance across all 16 scaffolds for two linear mixed models - the upper with no fixed effects and the lower with TSR as a covariate. Dashed black horizontal bar indicates the 0.05 FDR threshold, while the dashed grey bar above it indicates the Bonferroni correction threshold. X axis is scaled by the length of each scaffold. B) Manhattan plot for scaffold 5 (TSR model), which contains the target gene for glyphosate, EPSPS. The vertical grey rectangle indicates a previously identified ~6.5Mb EPSPS amplification present across individuals in this dataset. C) The proportion of significant SNPs present on each scaffold for each linear mixed model. D) The number of significant hits per gene in the TSR model.

The relative importance of monogenic versus polygenic mechanisms of resistance can be addressed by comparing the percent of variance explained by genome-wide SNPs in the TSR + Copy Number GWA (polygenic effects) to the r^2 of glyphosate resistance on TSR and EPSPS copy number (monogenic effects). Controlling for both EPSPS copy number and TSR leads to a marginal PVE estimate of 0.374, implying that genome-wide SNPs can explain 37% after accounting for resistance associated with EPSPS. Similarly, a more direct comparator, a GWA using the residuals of a regression of phenotypic resistance on TSR + EPSPS copy number (as opposed to a GWA model with TSR + EPSPS copy number as covariates) results in a PVE of

35%. Given that the initial multiple regression model of resistance on monogenic mechanisms (TSR + EPSPS copy number) results in an $r^2 = 0.33$ ($F_{2,152} = 38.69$, $p = 2.62e\text{-}14$), we can thus infer that 24% of the total variance in glyphosate resistance ((1-0.33) x 0.35) can be explained by genome-wide SNPs —implying a near equivalent importance of NTSR to monogenic mechanisms of resistance.

GO enrichment and candidate genes

The set of 196 significant SNPs with q-values < 0.05 in the TSR model corresponded to 125 unique *A. tuberculatus* genes. 70/125 of these genes were mappable to *A. thaliana* based on orthology, and thus could be used as input for enrichment analyses and candidate gene exploration. Only 5 genes had more than 4 associated hits; three encode unknown proteins and the other two encode tubulin-folding cofactor D and RING membrane-anchor 1 (**Fig. 2D**). A Fisher's exact test showed this set of genes to be significantly enriched for 17 hierarchical categories of PANTHER GO-Slim molecular function **(Table 1)**, however significance of these categories did not withstand multiple test correction. While we did not detect evidence for certain molecular functions or biological processes playing a particularly important role in glyphosate-resistance, our high quality set of resistance-related SNPs nonetheless shows functions consistent with our understanding of NTSR, which we present below.

Many of the molecular functions of genes in our set of significant SNP set were especially notable given previous work on particular gene families and the three known molecular phases of NTSR (Gaines et al., 2020): herbicide detoxification (via oxidation, hydrolysis, or reduction) followed by conjugation, and compartmentalization/ transport into the vacuole or extracellular space. Resistance-related alleles show molecular function in catalytic activity, hydrolase activity (via both lipase and hydrolases), and monooxygenase activity (i.e., cytochrome P450s) - potentially playing roles in phase I or II NTSR (**Table 2**). Cation channel activity, transmembrane transporter activity, and ion gated channel activity may be related to phase III, transport into the vacuole. In agreement, Panther GO-Slim cellular components implicate genes involved in the vacuolar membrane and endocytic vesicles (⅖ classes with raw p-values < 0.05) (although distinct from those genes involved in transport) **(Sup Table 2)**.

Table 2. Panther GO-Slim molecular function for 196 SNPs in 125 *A. tuberculatus* genes (70 *A. thaliana* orthologs) significantly associated with glyphosate resistance. Note that these classes are not significant after FDR correction.

Panther GO-Slim Molecular Function	Total	Found	Expected	Fold enrichment	Raw p-value	FDR
triglyceride lipase activity	1	1	0	> 100	5.1E-03	1.00
hydrolase activity	1631	9	4.2	2.2	3.6E-02	1.00
catalytic activity	4625	20	11.8	1.7	1.5E-02	7.5E-01
lipase activity	45	2	0.1	17.4	6.4E-03	4.8E-01
cation gated channel activity	19	2	0.12	16.32	7.24E-03	4.03E-01
passive transmembrane transporter activity	77	2	0.2	10.2	1.7E-02	7.73E-01
ion gated channel activity	40	2	0.1	19.6	5.2E-03	4.03E-01
DNA-binding transcription activator activity	41	2	0.1	19.1	5.4E-03	6.00E-01
hydrolase activity, acting on C-N bonds	79	2	0.2	9.9	1.8E-02	6.75E-01
tubulin binding	80	2	0.2	9.8	1.9E-02	6.38E-01
GTPase activity	145	3	0.4	8.1	6.4E-03	4.09E-01
monooxygenase activity	128	2	0.3	6.1	4.4E-02	1.00

These genes also function in diverse biological processes, including several catabolic and metabolic processes (glucose-6-phosphate, lipid, glycerol-lipid, aglycerol, neutral lipid, ammonium ion and triglyceride), and cell fate specification. While these terms may be intuitive given the phases of NTSR described above, unexpectedly, we also found genes functioning in leaf and flower development, regionalization, and circadian rhythm. These four terms relate to two genes: *YABBY1* and *GIGANTEA*. *GIGANTEA* is of particular interest, as it encodes a nuclear protein with pleiotropic roles in flowering time, stress response, circadian clock regulation, and there even being prior evidence for providing herbicide tolerance by increasing oxidative stress resistance (Kurepa, Smalle, Va, Montagu, & Inzé, 1998; Mishra & Panigrahi, 2015). Another example of a pleiotropically acting gene that has significant association with glyphosate resistance is *RAX1*, which affects shoot branching and in turn is transcriptionally regulated by a WRKY transcription factor, also encompassed in our significant SNP set (**Table 3**). Both *YABB1* and *GIGANTEA* genes (and associated molecular function categories) are absent from the significant SNP set of the covariate-free model, implying that the effect of the genes on glyphosate resistance changes depending on the TSR context—in other words, these genes may epistatically interact with the TSR mutation. The last example of pleiotropy in this set of putative-resistance genes hits closer to home—six genes that have been previously implicated as being either the direct target of, interacting with, or conferring resistance to other herbicidal compounds.

At the same time, some of our genes have putative prior evidence for a direct response to glyphosate exposure: several ERF genes (notably ERF105), a NADPH reductase, a glutamate receptor, and RMA1. Overall, we find many candidate genes are worth listing due to their previously identified roles in response to herbicides (Table 3). To be sure, previous associations of these genes with other herbicides does not prove a role in conferring glyphosate resistance in these populations. They do suggest, however, that these genes play a general role in plant xenobiotic responses or ameliorating general plant stresses.

Table 3. List of candidate NTSR to glyphosate genes of interest, significant after FDR correction and compared against evidence from the literature.

Encoded protein(s)	*A. thaliana* ID	Evidence from the literature	References
AMP deaminase	AT2G38280	Target of the herbicidal compound deaminoformycin	(Sabina, Paul, Ferl, Laber, & Lindell, 2007)
Cytochrome P450s: CYP77A4, CYP76C6, & CYP722A1	AT5G04660, AT1G33720, AT1G19630	Broad involvement of the cytochrome P450 gene family in NTSR in several species	(Gaines et al., 2020; Yuan et al., 2007)
Ethylene responsive transcription factors: ERF2, ERF061, ERF011, & ERF105	AT5G47220, AT1G64380, AT3G50260, AT5G51190	Expression of *ERF105* is induced in response to paraquat in *A. palmeri,* and repressed in response to glyphosate in *A. thaliana*	(Doğramacı et al., 2015; Illgen, Zintl, Zuther, Hincha, & Schmülling, 2020)
Ferredoxin, NADPH reductase	AT5G66190	- Part of the dicamba monooxygenase, an enzyme used to engineer resistance to the herbicide dicamba - Part of synthetic glyphosate acetyltransferase - Utilized by P450s to insert molecular oxygen in herbicides to make them more reactive or more soluble	(Green & Owen, 2011) (Bruggeman, Kuehler, & Weeks, 2014) (Gaines et al., 2020)
GDSL esterase/lipase	AT3G62280	Bioactivates an ACCase herbicide in blackgrass	(Ian Cummins & Edwards, 2004)
GIGANTEA	AT1G22770	Pleiotropic roles in flowering time, stress response, circadian clock regulation, and herbicide tolerance	(Kurepa et al., 1998; Mishra & Panigrahi, 2015)
Glutamate receptor	AT1G42540	Glutamate uptake altered by glyphosate	(Gomes et al., 2014; Serra et al., 2013)
Pectinesterase inhibitor 11	AT3G47380	Member of this gene family differentially expressed in chickpea treated with herbicide imidazoline	(Iquebal et al., 2017)
Ubiquitin-conjugating enzyme E2 32	AT3G17000	Mediates tolerance in *A. thaliana* to oxidative stress induced by paraquat herbicide	(Cui, Liu, Li, Yang, & Xie, 2012)
RAX1 Myb-like transcription factor	AT5G23000	Transcriptionally regulated by WRKY transcription factor, controls shoot branching	(D. Guo et al., 2015)
Transketolase-2	AT2G45290	Target of naturally synthesized herbicidal compound, α-Terthienyl	(B. Zhao, Huo, Liu, Zhang, & Dong, 2018)
E3 Ubiquitin-protein ligase RMA1	AT4G03510	A related E3 ligase mediates response to glyphosate in *A. thaliana*	(Faus et al., 2015) (Luo et al., 2016)
WRKY transcription factor	AT2G04880	- Based on expression data implicated in cross resistance in *A. thaliana* - Confers resistance to dicamba herbicide	(Mahmood, Mathiassen, Kristensen, & Kudsk, 2016) (Gleason, Foley, & Singh, 2011)
YABBY transcription factor	AT2G45190	Mediates stress responses and flower/leaf development	(Yang et al., 2018)

445

Genomic context of putative NTSR alleles

We posited that a closer look at the genomic context of the loci in our set of candidate genes would shed light on the strength of selection, genetic architecture, and heterogeneity of NTSR among populations. The effect sizes of our alleles are highly asymmetric in direction; only 4

SNPs have negative effects on glyphosate resistance compared to 192 positive effects, relative to susceptible reference, with this deficit being significant relative to to the null distribution (proportion of negative effect SNPs = 0.021, 95% CI of empirical null = 0.031, 0.082). Our significant SNPs tend to fall into two categories, rare alleles of modest effect and common alleles with small effects (**Fig. 3A)**. The negative relationship between effect size and frequency may be due to biological or methodological properties, or a combination of the two: resistance alleles with larger effects may have more deleterious pleiotropic effects (as in (Kryukov, Pennacchio, & Sunyaev, 2007; Marouli et al., 2017; Tennessen et al., 2012)), a lack of power makes it difficult to detect rare, small-effect alleles, and the Winner's Curse leads to an overestimation of detected effect sizes (D. J. Liu & Leal, 2012). Despite this apparent tradeoff, our set of resistance-related SNPs is enriched both for absolute effect sizes compared to the null distribution (median beta = 1.593185, 95% CIs of median empirical null: 1.371, 1.561), and for elevated allele frequencies (median AF = 0.097, 95% CIs of median empirical null: 0.061, 0.082) (**Fig. 3B).**

The set of resistance-related SNPs also show elevated values for the Weir-Cockerham F_{ST} diversity estimator between resistant and susceptible individuals, compared to both the genome-wide distribution and the null distribution (median F_{ST} = 0.099, 95% CIs of median empirical null: 0.063, 0.072) (**Fig. 3C)**. Notably, this pattern is independent of differences in the allele frequencies of SNP sets among the permuted and observed GWAS (**Fig 3C inset)**. This pattern also remains consistent after removing significant SNPs on the *EPSPS* containing scaffold 5, implying this excess is not driven by monogenic selection at *EPSPS*.

We also tested for signals of directional and balancing selection on glyphosate resistance-related alleles by examining the length and homozygosity of resistant versus susceptible haplotypes. We first investigated signals of extended haplotype homozygosity (EHH), a haplotype-based test that has been used to assay signals of positive selection around focal SNPs based on decay of homozygosity. Despite enrichment for allelic differentiation, we find no evidence of an enrichment of average EHH 10 kb around either side of observed resistant alleles compared to permuted (mean R allele EHH = 0.0703, 95% CIs of mean empirical null: 0.0577, 0.0737) (**Fig 3D)**, although we observed a slight excess of mean EHH for non-Scaffold 5 resistance alleles (mean R allele EHH = 0.0756). Taking the integral of the EHH by distance curve, referred to as

integrated Haplotype Homozygosity (iHH; (Voight, Kudaravalli, Wen, & Pritchard, 2006)) reveals that observed resistant haplotypes are smaller than expected under the empirical null distribution (median iHH = 0.00405724, 95% CIs of median empirical null: 0.0059, 0.0084) (**Fig 3E**; difficult to visualize in Fig 3D as this pattern is in part driven by larger variance in the EHH curve among permuted SNPs**)**. Finally, the standardized integrated Haplotype Score (iHS) (which compares differences in iHH between reference and ancestral alleles, standardized for allele frequency; (Voight et al., 2006)) shows enrichment for unusually small homozygous tracts of resistant relative to susceptible haplotypes, given differences in their allele frequencies (median standardized iHS = -0.9544, 95% CIs of median empirical null: -0.3665, -0.1214) (also the case for non-Scaffold 5 SNPs) (**Fig F).**

Together the strong allelic-differentiation and short-haplotypes seem to suggest a role for balancing selection, and so as a final check we also interrogated 0-fold and 4-fold diversity and Tajima's D in 10kb windows surrounding focal SNPs which is also expected to be elevated under balancing selection. While we see evidence for elevated 0-fold diversity and Tajima's D surrounding resistance-related SNPs genome-wide (mean 0-fold π = 0.0031, [95% CIs: 0.000, 0.0013]; median 0-fold Tajima's D = -0.502 [95% CIs:-0.608, -0.518]), we see a much more muted signal for 4-fold sites (mean 4-fold π =0.0075, [95% CIs: 0.0020, 0.0057]; median 4-fold Tajima's D = -0.399 [95% CIs: -0.482, -0.359]), that appears to be in part driven by scaffold 5 (mean 4-fold π at non-scaf 5 SNPs = 0.0032, [95% CIs: 0.0015, 0.0065]) (**Sup Figure 1**).

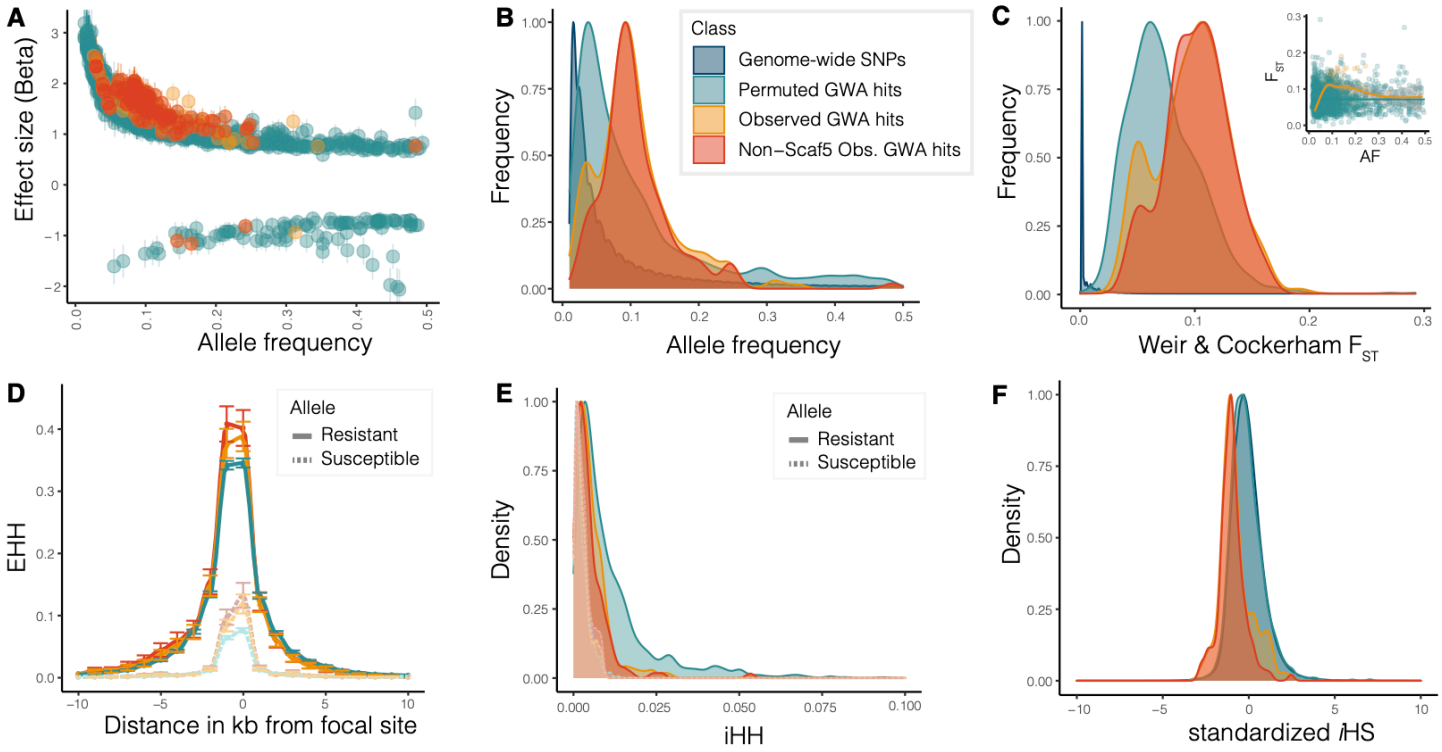


Fig 3. Genomic summaries of SNPs significantly associated with glyphosate resistance (for all significant SNPs, and non-Scaffold 5 significant hits) compared to the empirical null distribution and genome-wide background. A) Effect sizes relative to allele frequencies, B) the distribution of allele frequencies, C) the distribution of F_{st} values (inset showing association between F_{st} and allele frequency), D) expected haplotype homozygosity (EHH) around focal resistance-related and susceptibility-related alleles, E) the integrated haplotype homozygosity (iHH) of resistant-related versus susceptibility-related haplotypes, F) standardized integrated haplotype score (iHS), treating resistant alleles as the derived allele (note a positive value indicates an excess of derived relative to ancestral haplotype size, as in (Szpiech & Hernandez, 2014)). The legend in the bottom right corner applies to all plots, while legends for D and E additionally distinguish allele type.

Discussion

Herbicides, when appropriately applied in agricultural settings, are often lethal, and for this reason they elicit very strong and rapid evolutionary responses in weed populations. The most commonly recognized genetic responses to these strong selection pressures are the appearance of mutations at the loci whose products are directly targeted by the herbicide. How often selection by herbicides leads to easily detectable changes at other loci in the genome has remained unclear. In this study, we investigated the genetic architecture of resistance to glyphosate herbicides in a

problematic agricultural weed, *Amaranthus tuberculatus.* As proof of concept, GWA correctly identified *EPSPS,* the target gene of glyphosate herbicides, as significantly associated with resistance. Consistent with polygenic NTSR being widespread across populations and the genome, however, we find that in addition SNPs in more than 100 distinct genes across all 16 chromosomes are also associated. While we refer to these significantly associated alleles as resistance-related, it is important to note that the genetic architecture we characterize is likely to be the product of a complex history of selection and accumulation of background-specific compensatory mutations (as in (Craig MacLean, Hall, Perron, & Buckling, 2010)). In particular, nearly all of our resistant individuals have EPSPS amplification variants and/or TSR, so many of our identified resistance-associated SNPs may be epistatically related to stress-tolerance, conditional on the presence of glyphosate resistance. Compared to both the genome-wide background and null-expectation, we find that glyphosate resistance-related alleles are enriched for intermediate frequencies and large effects. Together, these results fit a scenario of a genetic architecture of NTSR shaped by strong selection, with common alleles shared across populations and rare, population-specific and possibly background-dependent alleles. Despite the recognition that positive selection from herbicides should be rampant in the field, nearly all of our population genomic analyses point to balancing selection maintaining resistance and susceptibility related alleles, likely over longer evolutionary timescales than herbicide usage. Below, we discuss the genetic architecture of glyphosate resistance, the role of selection in shaping it, and candidate NTSR genes.

Model comparison

With previous knowledge of two genetic causes underlying glyphosate resistance—a non-synonymous mutation within *EPSPS* that prevents glyphosate from inhibiting the encoded enzyme, and amplification of *EPSPS* that overcomes the inhibitory effect of glyphosate through overproduction of the *EPSPS* protein—we sought to understand how these covariates influence the inferred genetic architecture of resistance. We compared linear mixed effect models, with iterative additions of these covariates, based on metrics of GWA model-fitting quality (**Table 1**). Our best model in terms of both inflation, percent of residual phenotypic variation explained, and FDR was one that only included TSR as a covariate. Reassuringly, this model showed 32 SNPs within the EPSPS-amplification significantly related to glyphosate resistance, genotypes which

together explain 76% of the variance in EPSPS copy number (based on multiple linear regression), indicating that the majority of the variance in resistance due to the amplification has been accounted for. However, it is worth keeping in mind that some significant SNPs in our set from the TSR model may be indirectly interacting with resistance through their association (biological or statistical) with EPSPS.

Genome-wide SNPs explain 35% of residual variance in glyphosate resistance (PVE based on residual resistance values after removing variance explained by monogenic modes), after accounting for the 33% of variance in resistance explained by monogenic mechanisms (as inferred from multiple regression). Given that the residual variation in resistance should approximate 1-(monogenic r^2), genome-wide SNPs in our model can thus explain 24% of total variance in resistance—implying NTSR mechanisms can explain nearly as much variance in phenotypic resistance as monogenic mechanisms.

Effect size & Allele frequencies

The effect sizes of SNPs associated with resistance tend to be negatively correlated with allele frequency (**Fig 2A)**. The lack of rare alleles with small effects is likely to be driven by lack of statistical power—a key driver in the problem of missing heritability (Manolio et al., 2009). For human height, a classic example of a polygenic trait where small effect sizes should be prevalent, only 83 significantly associated alleles with a frequency < 0.05 could be identified with a sample size of 711,428 individuals, and their effect sizes were >10 times that of common variants on average (Marouli et al., 2017). The median effect size (beta) of rare alleles (<0.05) in our GWAS was 2.34, while for common alleles (=>0.05) it was only 1.52, showing an excess of large-effect common alleles, when compared to the extreme human height case. The excess of large effect, common alleles is also consistent with our expectation of resistance as a much less polygenic trait than height. These beta values can be interpreted as the effect of increasing the count of the allele by 1, per unit of resistance. In our case, a unit of resistance on a scale of 1-5 corresponds roughly to a 20% decrease in percent damage from glyphosate application. Thus, for every increase in allele count, the median effect size rare allele may lead to a near 50% decrease in percent damage.

Aside from power, the negative correlation between allele frequency and effect size and the general lack of common, large effect alleles may be a consequence of selection. For human diseases and *Capsella grandiflora* gene expression, this phenomenon has been explained by greater deleterious, and possibly pleiotropic, costs of large effect alleles, such that purifying selection reduces their frequency within populations (Eyre-Walker, 2010; Josephs, Lee, Stinchcombe, & Wright, 2015; Kryukov et al., 2007; Marouli et al., 2017). While we see this negative correlation in our observed SNP set, both large effects (**Fig 2A)** and common alleles **(Fig 2B)** are enriched relative to the null expectation. The enrichment for large effect sizes is likely to reflect particularly strong selection in shaping the genetic architecture for herbicide resistance (Eyre-Walker, 2010), although that does not explain the additional enrichment for common variants. However, the population genetic framework in (Eyre-Walker, 2010) models a polygenic trait's influence on fitness through its pleiotropic costs, where a tradeoff between large effect sizes and rare allele frequencies mediates these costs. With herbicide resistance, selection for resistance approaches near lethality (s ~ 1) and the benefits of large-effect, common alleles are likely to outweigh any pleiotropic costs (at least in the short term). Thus, as opposed to most polygenic traits where the cumulative effect of rare large-effect and common small-effect alleles explain most phenotypic variance, for herbicide resistance, populations additionally respond to selection through increasing the frequencies of large effect variants.

The range in allele frequency and enrichment of high frequency alleles also implies that a subset of resistance-alleles are shared among populations, whereas others are population specific. Rare-population specific alleles are likely to be the work of compensatory mutations arising on already resistant backgrounds (as is the case for antibiotic resistance, summarized in (Craig MacLean et al., 2010)), that interact epistatically to strengthen resistance to herbicides or mediate their costs in the absence of herbicides.

Candidate genes

We found 196 resistance-associated SNPs in the TSR model, corresponding to 125 unique *Amaranthus tuberculatus* genes. Of these genes, we could annotate 70 using orthology-based approaches. We found our gene set had diverse roles in molecular function, biological processes, and cellular components, with various terms in each significantly enriched, but not after multiple test correction (we focus our discussion on molecular function, as it is the most relevant in the

weed science literature on mechanisms of non-target site resistance (Gaines et al., 2020)). While not significantly enriched after multiple test correction, molecular functions in our gene set was highly relevant to the three-phase schema of non-target site resistance (Gaines et al., 2020)—detoxification, conjugation, and transport—including substantial monooxygenase function by paradigmatic cytochrome P450s. Thus, while we did not detect evidence of a particular molecular function playing a foremost role in glyphosate resistance, we do have strong evidence for the involvement of our particular genes and their relevance to NTSR, which we discuss further below.

Several types of proteins encoded by genes in our significant set have been previously linked to resistance to glyphosate or are a part of the shikimate pathway targeted by glyphosate: Ferredoxin NADPH reductase belongs to two synthetic enzymes (dicamba monooxygenase and glyphosate acetyltransferase) that are used to produce genetically modified herbicide resistant plants (Bruggeman et al., 2014; Gaines et al., 2020; Green & Owen, 2011); ERF0105 is differentially expressed in glufosinate treated *A. thaliana* (Salas-Perez et al., 2018); uridine kinase-like protein 3 mediates responses to glyphosate in *A. thaliana* (Faus et al., 2015; Luo et al., 2016); and glutamate receptors may be important, because glutamate uptake is altered by glyphosate (Gomes et al., 2014; Serra et al., 2013). In our set of genes, these may be those most likely to directly influence resistance to glyphosate herbicides, although it is statistically impossible to verify the strength of evidence for each of them. Instead, they may be interesting candidate NTSR genes for further functional validation.

Many proteins encoded by genes tagged by our set of significant SNPs also appear to be the target of other herbicides or herbicidal compounds. For example: Transketolase-2 is the target of a naturally derived herbicidal compound α-Terthienyl (Sabina et al., 2007), GDSL esterase/lipase bioactivates an ACCase herbicide in blackgrass (Ian Cummins & Edwards, 2004), ubiquitin-conjugating enzyme E2 32 is induced by the herbicide paraquat in *A. thaliana* (Cui et al., 2012), and AMP deaminase is the target of the herbicidal compound deaminoformycin (Sabina et al., 2007). That a GWA for glyphosate resistance results in several significant associations with these genes implies that they may pleiotropically function to confer cross-resistance to several modes of herbicides, or that glyphosate resistance is associated with other, unassayed modes of

resistance in our samples. With respect to the former hypothesis, we found one gene, encoding a WRKY transcription factor, that has been previously shown to contribute to herbicide cross-resistance in *A. thaliana* based on expression data.

Genes in our associated set also seem to have pleiotropic effects on non-resistance related traits—*GIGANTEA* has been previously identified as playing roles in flowering time, stress response, circadian clock regulation and herbicide tolerance, while *YABBY* genes function in stress response and leaf/flower development and RAX1 in shoot branching. All in all, we identify a diverse set of candidate genes, with previously identified roles in resistance to many herbicides, including glyphosate, and pleiotropic roles in cross resistance, life history and physiological traits.

While pleiotropy may be one explanation for why we see genes with diverse functions in our GWA for glyphosate resistance, a complex history of compounded selection from various herbicides and related shifts in life history optima could also drive this pattern. The problem of correlated traits is endemic to all GWA style analyses (Coop, 2019; Novembre & Barton, 2018; Racimo, Berg, & Pickrell, 2018), which is not a new problem in evolutionary genetics (Lande, 1979; Lande & Arnold, 1983). GWAS are inherently correlative, with traits (response variables) related to SNP genotypes through a statistical model. It is important to keep in mind that these types of analyses, in addition to describing the genetic basis of the response variable, can also capture correlated traits in at least two ways. First, after the origin of TSR resistance or copy number expansion, it is likely that any trait that improves plant performance in the presence of glyphosate damage (e.g., growth rates, generalized stress responses, changes to photosynthetic performance, water relations, phenology, etc) will be favored, even if these traits are physiologically and pleiotropically unrelated to detoxification, conjugation, or transport of herbicides. Consequently, SNPs that improve these traits will be associated with phenotypic assays of resistance and observed as GWAS hits. Second, any agricultural practice (pesticide use, manuring, tilling, or irrigation), soil characteristic, or climate variable that is correlated with glyphosate resistance will create a set of co-selected traits (e.g., resistance, nutrient uptake, seed germination dormancy, flowering time, drought tolerance), and these may also be detected as a

GWAS hit for herbicide resistance, even if their function is for other ecologically important traits.

That being said, at least part of this signal we describe here is likely to be independent from correlated selection, especially given that we see evidence of selection on *EPSPS,* a known target of glyphosate herbicides. Furthermore, interpreting the function of certain genes in light of the historical context of field sources of selection, implies that three links to other herbicidal compounds are unlikely to be driven by correlated selection from other herbicides and rather may be linked through stress-response; ACCase herbicides have no activity in *Amaranthus tuberculatus*, and deaminoformycin & α-terthienyl are not used in the field for amaranth control. Nonetheless, the interpretation of these resistance-related alleles as strictly driven by selection only from glyphosate herbicides is incorrect—these resistance-related alleles are likely to characterize a complex history of agricultural selection in resistant weeds. To disentangle the complexities of how shared the genetic basis of NTSR is in response to different herbicides and with other non-resistance related traits, complementary quantitative genetic and population genomic methods could be applied to cohort of genotypes with many phenotyped traits to estimate correlated selection and corresponding genetic architectures.

Differentiation, haplotype homozygosity, and a role for balancing selection

Our set of resistance-related alleles are not only enriched for effect sizes and high allele frequencies, but also for genetic differentiation (Weir & Cockerhams Fst) and in unexpected ways with respect to haplotype length and homozygosity (**Fig 3C-F)**. Enrichment relative to the permuted expectation is likely to reflect signals of selection, as it is above and beyond what would be expected given the population structure in our samples for 100s of random traits. While by definition, GWAS are expected to find alleles with high allele frequency differences and differentiation for a trait of interest, our resistance alleles are enriched even beyond permuted null expectations (**Fig 3C)**—consistent with a role of strong selection in shaping allele frequency differences among susceptible and resistant individuals. It should be noted that this signal is independent of differences in allele frequencies between our permuted and observed SNP set; F_{st} in our permuted SNPs shows no correlation with allele frequency (but higher variance at lower

frequencies), whereas Fst in our observed SNP set shows no linear correlation but an enrichment of Fst at intermediate allele frequencies (**Fig 3C inset)**.

Selection on genome-wide resistance-related alleles may be in the form of positive and/or balancing selection. Because net selection for and against herbicide resistance should vary depending on environment, assuming a direct or pleiotropic cost of resistance, both susceptible and resistant individuals may be alternately favored in herbicide and herbicide-free settings over many generations. Weed science research has long been interested in the costs of herbicide resistance mutations, and much work on the topic has led to mixed results, with fitness costs dependent on the mutation, organism, and environment (summarised in (Baucom, 2019)). Nonetheless, studies on the costs of resistance in the absence of herbicides have focused primarily on TSR mechanisms, and pleiotropic costs of resistance as a trait should be much more ubiquitous where hundreds of diverse genes are involved (e.g. 90% of all human GWAS hits overlap multiple traits; (Watanabe et al., 2019)).

We find no evidence that extended haplotype homozygosity (EHH) is enriched in a distance of 10kb around significant hits in our observed GWA compared to the null expectation and alternative susceptible haplotypes (**Fig 3D)**. However, we see both a significant deficit of integrated haplotype homozygosity for the resistance haplotype (**Fig 3E)** and the standardized integrated haplotype score, relative to the permuted null expectation **(Fig 3F)**. Since iHS is standardized by genome-wide empirical allele frequency distributions, this can be interpreted as unusually fast EHH decay, or unusually small homozygous haplotypes, associated with resistant relative to susceptible alleles (**Fig 3E)** (Voight et al., 2006). Particularly short haplotypes associated with resistance are not expected under partial selective sweeps, which should increase the homozygous haplotype length of selected compared to unselected alleles (Voight et al., 2006)). In contrast, balancing selection, which is expected to maintain diversity over longer timescales, should lead to particularly short tracts of homozygous haplotypes as old haplotypes have both high diversity and allelic associations broken-up overtime due to more opportunity for recombination (B. Charlesworth, Nordborg, & Charlesworth, 1997; D. Charlesworth, 2006; Hudson & Kaplan, 1988). A significant deficit of standardized iHSs & resistant iHH values implies that resistance alleles have been maintained over longer timescales than expected.

Under the long-term maintenance of genetic diversity—or balancing selection—we also expect to see an excess of neutral (4 fold) genetic diversity in genomic windows surrounding focal alleles, and an excess of positive Tajima's D. While we see enrichment of these statistics in surrounding 0 fold sites, we do not see this for 4-fold sites. One explanation for this lack of signal may be due to the tendency of our resistance-related haplotypes to be particularly short, leaving us with little chance to detect long range signals of selection in linked neutral diversity. Therefore, while particularly high resistance-related allele frequencies and differentiation seemed to initially imply directional positive selection, taken together with haplotype-based inference, these results are most consistent with the maintenance of resistance-related alleles through balancing selection.

That herbicide-resistance alleles appear to be under balancing selection seems like a contradiction of timescales—selection from herbicides should be extremely recent (on the scale of decades), while detectable signals of balancing selection are typically very old (on the scale of millions of years) (D. Charlesworth, 2006). Rather than recent fluctuations in strong selection due to alternating herbicide use, unusually short resistant haplotypes may be possible if these alleles represent a subset of broadly functioning stress-response genes in which diversity has been maintained in populations over longer evolutionary time, potentially due to fluctuating selection pressures on stress-related traits over space and time. This further highlights the potential for pre-existing standing genetic variation on resistance-associated variation prior to herbicide usage.

Early seminal population genetic theory on herbicide resistance posits that even with rotating crop and herbicide use (a rotation scenario of one year on/one year off)—with selection coefficients (*s)* for resistance ranging between 0.75 and 0.99 within generations, and with a cost of resistance in the absence of herbicides as s=0.25—that a rare monogenic dominant allele will reach a frequency of ~0.7 (s=0.75) or go to fixation (s=0.99) within 20 years (Jasieniuk, Anita L. Brûlé-Babel, & Ian N. Morrison, 1996). However, our results illustrate that the genetic architecture of resistance is more complex and polygenic than a single dominant allele, encompassing a range of allele frequencies and effect sizes. This complex basis means that small

to moderate changes in the frequency of many alleles can appreciably alter phenotypic values, with weaker selection on each allele individually. Compounded with the possible role of long-term balancing that we find here, diversity in polygenic resistance-related alleles may persist in populations over much longer timescales then initially hypothesized.

Conclusion

In conclusion, our study identifies 125 genes across all 16 chromosomes associated with glyphosate resistance. These alleles range in effect size and allele frequency while being enriched for high values of both, implying a drastic response of genome-wide alleles to selection from herbicides, but an in part heterogeneous architecture of resistance across populations. Consistent with the literature, we find resistance-related genes function in detoxification, conjugation, and transport, but we also find that these alleles may have pleiotropic roles in resistance to several classes of herbicides, and key life history traits. We find evidence of elevated differentiation, allele frequencies, and unusually short resistant haplotypes, implying a role of balancing selection in the maintenance of resistance-related alleles. Given that genome-wide SNPs can explain 25% of the total variance in glyphosate resistance in this sample of individuals, this work implies that we can remain optimistic about the prospects for phenotypic prediction of resistance in weed populations. Moreover, our results suggest that selection from herbicides may have more widespread consequences on genomic diversity than initially assumed. Further work to functionally validate our candidate genes will shed light on mechanism and consistency of these alleles in conferring resistance across populations.

Author Contributions

JMK, SIW, and JRS designed the research; JMK conceived, analyzed, and wrote the paper; and JMK, SIW, JRS, PJT, and DW contributed feedback and edited the paper.

Acknowledgements

We would like to thank Tyler Kent, Amardeep Singh, and George Sandler for helpful and fun discussions. JK was supported by a Society for the Study of Evolution Rosemary Grant and an NSERC PGS-D, JRS and SIW were supported by Discovery Grants from NSERC Canada, SIW was additionally supported by a Canada Research Chair in Population Genomics, and DW was

supported by the Max Planck Society and Ministry of Science, Research and the Arts of Baden-Württemberg in the Regio-Research-Alliance "Yield Stability in Dynamic Environments".

References

Baucom, R. S. (2019). Evolutionary and ecological insights from herbicide-resistant weeds: what have we learned about plant adaptation, and what is left to uncover? *The New Phytologist*, *223*(1), 68–82.

Bell, M. S., Hager, A. G., & Tranel, P. J. (2013). Multiple Resistance to Herbicides from Four Site-of-Action Groups in Waterhemp (Amaranthus tuberculatus). *Weed Science*, *61*(3), 460–468.

Benevenuto, J., Bhakta, M., Lohr, D. A., Ferrão, L. F. V., Resende, M. F. R., Kirst, M., … Munoz, P. (2019). Cost-effective detection of genome-wide signatures for 2,4-D herbicide resistance adaptation in red clover. *Scientific Reports*, Vol. 9. doi: 10.1038/s41598-019-55676-9

Bernards, M. L., Crespo, R. J., Kruger, G. R., Gaussoin, R., & Tranel, P. J. (2012). A Waterhemp (Amaranthus tuberculatus) Population Resistant to 2,4-D. *Weed Science*, *60*(3), 379–384.

Bock, D. G., Kantar, M. B., Caseys, C., Matthey-Doret, R., & Rieseberg, L. H. (2018). Evolution of invasiveness by genetic accommodation. *Nature Ecology & Evolution*, Vol. 2, pp. 991–999. doi: 10.1038/s41559-018-0553-z

Bruggeman, A. J., Kuehler, D., & Weeks, D. P. (2014). Evaluation of three herbicide resistance genes for use in genetic transformations and for potential crop protection in algae production. *Plant Biotechnology Journal*, *12*(7), 894–902.

Busi, R., Porri, A., Gaines, T. A., & Powles, S. B. (2018). Pyroxasulfone resistance in Lolium rigidum is metabolism-based. *Pesticide Biochemistry and Physiology*, *148*, 74–80.

Charlesworth, B., Nordborg, M., & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research*, *70*(2), 155–174.

Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics*, *2*(4), e64.

Coop, G. (2019). Reading tea leaves? Polygenic scores and differences in traits among groups. Retrieved from http://arxiv.org/abs/1909.00892

Craig MacLean, R., Hall, A. R., Perron, G. G., & Buckling, A. (2010). The population genetics of antibiotic resistance: integrating molecular mechanisms and treatment contexts. *Nature Reviews. Genetics*, *11*(6), 405–414.

Cui, F., Liu, L., Li, Q., Yang, C., & Xie, Q. (2012). UBC32 mediated oxidative tolerance in Arabidopsis. *Journal of Genetics and Genomics = Yi Chuan Xue Bao*, *39*(8), 415–417.

Cummins, I., Cole, D. J., & Edwards, R. (1999). A role for glutathione transferases functioning as glutathione peroxidases in resistance to multiple herbicides in black-grass. *The Plant Journal: For Cell and Molecular Biology*, *18*(3), 285–292.

Cummins, I., & Edwards, R. (2004). Purification and cloning of an esterase from the weed black-grass (Alopecurus myosuroides), which bioactivates aryloxyphenoxypropionate herbicides. *The Plant Journal*, Vol. 39, pp. 894–904. doi: 10.1111/j.1365-313x.2004.02174.x

Cummins, I., Wortley, D. J., Sabbadin, F., He, Z., Coxon, C. R., Straker, H. E., … Edwards, R. (2013). Key role for a glutathione transferase in multiple-herbicide resistance in grass

weeds. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(15), 5812–5817.

Delaneau, O., Zagury, J.-F., & Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. *Nature Methods*, *10*(1), 5–6.

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

Delye, C., Gardin, J. A. C., Boucansaud, K., Chauvel, B., & Petit, C. (2011). Non-target-site-based resistance should be the centre of attention for herbicide resistance research: Alopecurus myosuroides as an illustration. *Weed Research*, *51*(5), 433–437.

Doğramacı, M., Foley, M. E., Horvath, D. P., Hernandez, A. G., Khetani, R. S., Fields, C. J., … Anderson, J. V. (2015). Glyphosate's impact on vegetative growth in leafy spurge identifies molecular processes and hormone cross-talk associated with increased branching. *BMC Genomics*, Vol. 16. doi: 10.1186/s12864-015-1627-9

Duhoux, A., Carrère, S., Duhoux, A., & Délye, C. (2017). Transcriptional markers enable identification of rye-grass (Lolium sp.) plants with non-target-site-based resistance to herbicides inhibiting acetolactate-synthase. *Plant Science*, Vol. 257, pp. 22–36. doi: 10.1016/j.plantsci.2017.01.009

Duhoux, A., Carrère, S., Gouzy, J., Bonin, L., & Délye, C. (2015). RNA-Seq analysis of rye-grass transcriptomic response to an herbicide inhibiting acetolactate-synthase identifies transcripts linked to non-target-site-based resistance. *Plant Molecular Biology*, Vol. 87, pp. 473–487. doi: 10.1007/s11103-015-0292-3

Duke, S. O., & Powles, S. B. (2008). Glyphosate: a once-in-a-century herbicide. *Pest Management Science: Formerly Pesticide Science*, *64*(4), 319–325.

Emms, D. M., & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, *16*, 157.

Epstein, B., Abou-Shanab, R. A. I., Shamseldin, A., Taylor, M. R., Guhlin, J., Burghardt, L. T., … Tiffin, P. (2018). Genome-Wide Association Analyses in the Model Rhizobium Ensifer meliloti. *mSphere*, *3*(5). doi: 10.1128/mSphere.00386-18

Exposito-Alonso, M., 500 Genomes Field Experiment Team, Burbano, H. A., Bossdorf, O., Nielsen, R., & Weigel, D. (2019). Natural selection on the Arabidopsis thaliana genome in present and future climates. *Nature*. doi: 10.1038/s41586-019-1520-9

Exposito-Alonso, M., Vasseur, F., Ding, W., Wang, G., Burbano, H. A., & Weigel, D. (2018). Genomic basis and evolutionary potential for extreme drought adaptation in Arabidopsis thaliana. *Nature Ecology & Evolution*, Vol. 2, pp. 352–358. doi: 10.1038/s41559-017-0423-0

Eyre-Walker, A. (2010). Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(suppl 1), 1752–1756.

Faus, I., Zabalza, A., Santiago, J., Nebauer, S. G., Royuela, M., Serrano, R., & Gadea, J. (2015). Protein kinase GCN2 mediates responses to glyphosate in Arabidopsis. *BMC Plant Biology*, *15*, 14.

Flood, P. J., van Heerwaarden, J., Becker, F., de Snoo, C. B., Harbinson, J., & Aarts, M. G. M. (2016). Whole-Genome Hitchhiking on an Organelle Mutation. *Current Biology: CB*, *26*(10), 1306–1311.

Gaines, T. A., Duke, S. O., Morran, S., Rigon, C. A. G., Tranel, P. J., Küpper, A., & Dayan, F. E. (2020). Mechanisms of evolved herbicide resistance. *The Journal of Biological Chemistry*. doi: 10.1074/jbc.REV120.013572

Gleason, C., Foley, R. C., & Singh, K. B. (2011). Mutant Analysis in Arabidopsis Provides Insight into the Molecular Mode of Action of the Auxinic Herbicide Dicamba. *PLoS ONE*, Vol. 6, p. e17245. doi: 10.1371/journal.pone.0017245

Gomes, M. P., Smedbol, E., Chalifour, A., Hénault-Ethier, L., Labrecque, M., Lepage, L., … Juneau, P. (2014). Alteration of plant physiology by glyphosate and its by-product aminomethylphosphonic acid: an overview. *Journal of Experimental Botany*, *65*(17), 4691–4703.

Green, J. M., & Owen, M. D. K. (2011). Herbicide-resistant crops: utilities and limitations for herbicide-resistant weed management. *Journal of Agricultural and Food Chemistry*, *59*(11), 5819–5829.

Guo, D., Zhang, J., Wang, X., Han, X., Wei, B., Wang, J., … Qin, G. (2015). The WRKY Transcription Factor WRKY71/EXB1 Controls Shoot Branching by Transcriptionally Regulating RAX Genes in Arabidopsis. *The Plant Cell*, *27*(11), 3112–3127.

Guo, J., Riggins, C. W., Hausman, N. E., Hager, A. G., Riechers, D. E., Davis, A. S., & Tranel, P. J. (2015). Nontarget-Site Resistance to ALS Inhibitors in Waterhemp (Amaranthus tuberculatus). *Weed Science*, *63*(2), 399–407.

Heap, I. (2010). The international herbicide-resistant weed database. *Available At:(accessed 4 October 2019)*.

Holliday, R. J., & Putwain, P. D. (1980). Evolution of Herbicide Resistance in Senecio vulgaris: Variation in Susceptibility to Simazine Between and Within Populations. *The Journal of Applied Ecology*, *17*(3), 779–791.

Hudson, R. R., & Kaplan, N. L. (1988). The coalescent process in models with selection and recombination. *Genetics*, *120*(3), 831–840.

Illgen, S., Zintl, S., Zuther, E., Hincha, D. K., & Schmülling, T. (2020). Characterisation of the ERF102 to ERF105 genes of Arabidopsis thaliana and their role in the response to cold stress. *Plant Molecular Biology*, *103*(3), 303–320.

Iquebal, M. A., Soren, K. R., Gangwar, P., Shanmugavadivel, P. S., Aravind, K., Singla, D., … Kumar, D. (2017). Discovery of Putative Herbicide Resistance Genes and Its Regulatory Network in Chickpea Using Transcriptome Sequencing. *Frontiers in Plant Science*, *8*, 958.

Jasieniuk, M., Anita L. Brûlé-Babel, & Ian N. Morrison. (1996). The Evolution and Genetics of Herbicide Resistance in Weeds. *Weed Science*, *44*(1), 176–193.

Johnson, R. C., Nelson, G. W., Troyer, J. L., Lautenberger, J. A., Kessing, B. D., Winkler, C. A., & O'Brien, S. J. (2010). Accounting for multiple comparisons in a genome-wide association study (GWAS). *BMC Genomics*, *11*(1). doi: 10.1186/1471-2164-11-724

Johnson, W. G., Davis, V. M., Kruger, G. R., & Weller, S. C. (2009). Influence of glyphosate-resistant cropping systems on weed species shifts and glyphosate-resistant weed populations. *European Journal of Agronomy: The Journal of the European Society for Agronomy*, *31*(3), 162–172.

Josephs, E. B., Lee, Y. W., Stinchcombe, J. R., & Wright, S. I. (2015). Association mapping reveals the role of purifying selection in the maintenance of genomic variation in gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(50), 15390–15395.

Kreiner, Giacomini, Bemm, Waithaka, Regalado, Lanz, Hildebrandt, Sikkema, Tranel, Weigel,

Stinchcombe, Wright. (2019a). Amaranthus tuberculatus PNAS 2019 Reference Genome [Data set]. *Multiple Modes of Convergent Adaptation in the Spread of Glyphosate-Resistant Amaranthus Tuberculatus*. Retrieved from https://genomevolution.org/coge/GenomeInfo.pl?gid=54057

Kreiner, Giacomini, Bemm, Waithaka, Regalado, Lanz, Hildebrandt, Sikkema, Tranel, Weigel, Stinchcombe, Wright. (2019b). Amaranthus tuberculatus PNAS 2019 Sequence Data [Data set]. *Multiple Modes of Convergent Adaptation in the Spread of Glyphosate-Resistant Amaranthus Tuberculatus*. Retrieved from https://www.ebi.ac.uk/ena/browser/view/PRJEB31711

Kreiner, J. M., Giacomini, D. A., Bemm, F., Waithaka, B., Regalado, J., Lanz, C., … Wright, S. I. (2019). Multiple modes of convergent adaptation in the spread of glyphosate-resistant Amaranthus tuberculatus. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(42), 21076–21084.

Kreiner, J. M., Stinchcombe, J. R., & Wright, S. I. (2018). Population Genomics of Herbicide Resistance: Adaptation via Evolutionary Rescue. *Annual Review of Plant Biology*, *69*, 611–635.

Kryukov, G. V., Pennacchio, L. A., & Sunyaev, S. R. (2007). Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *American Journal of Human Genetics*, *80*(4), 727–739.

Küpper, A., Manmathan, H. K., Giacomini, D., Patterson, E. L., McCloskey, W. B., & Gaines, T. A. (2018). Population Genetic Structure in Glyphosate-Resistant and -Susceptible Palmer Amaranth (Amaranthus palmeri) Populations Using Genotyping-by-sequencing (GBS). *Frontiers in Plant Science*, *9*, 29.

Küpper, A., Peter, F., Zöllner, P., Lorentz, L., Tranel, P. J., Beffa, R., & Gaines, T. A. (2018). Tembotrione detoxification in 4-hydroxyphenylpyruvate dioxygenase (HPPD) inhibitor-resistant Palmer amaranth (Amaranthus palmeri S. Wats.). *Pest Management Science*, *74*(10), 2325–2334.

Kurepa, J., Smalle, J., Va, M., Montagu, N., & Inzé, D. (1998). Oxidative stress tolerance and longevity in Arabidopsis: the late-flowering mutant gigantea is tolerant to paraquat. *The Plant Journal: For Cell and Molecular Biology*, *14*(6), 759–764.

Lande, R. (1979). QUANTITATIVE GENETIC ANALYSIS OF MULTIVARIATE EVOLUTION, APPLIED TO BRAIN:BODY SIZE ALLOMETRY. *Evolution; International Journal of Organic Evolution*, *33*(1Part2), 402–416.

Lande, R., & Arnold, S. J. (1983). The Measurement of Selection on Correlated Characters. *Evolution; International Journal of Organic Evolution*, *37*(6), 1210–1226.

Lee, R. M., Thimmapuram, J., Thinglum, K. A., Gong, G., Hernandez, A. G., Wright, C. L., … Tranel, P. J. (2009). Sampling the Waterhemp (Amaranthus tuberculatus) Genome Using Pyrosequencing Technology. *Weed Science*, *57*(5), 463–469.

Leslie, T., & Baucom, R. S. (2014). De Novo Assembly and Annotation of the Transcriptome of the Agricultural Weed Ipomoea purpurea Uncovers Gene Expression Changes Associated with Herbicide Resistance. *G3: Genes, Genomes, Genetics*, *4*(10), 2035–2047.

Liu, D. J., & Leal, S. M. (2012). Estimating genetic effects and quantifying missing heritability explained by identified rare-variant associations. *American Journal of Human Genetics*, *91*(4), 585–596.

Liu, W., Bai, S., Zhao, N., Jia, S., Li, W., Zhang, L., & Wang, J. (2018). Non-target site-based resistance to tribenuron-methyl and essential involved genes in Myosoton aquaticum (L.).

BMC Plant Biology, *18*(1), 225.

Luo, C., Cai, X.-T., Du, J., Zhao, T.-L., Wang, P.-F., Zhao, P.-X., … Xiang, C.-B. (2016). PARAQUAT TOLERANCE3 Is an E3 Ligase That Switches off Activated Oxidative Response by Targeting Histone-Modifying PROTEIN METHYLTRANSFERASE4b. *PLOS Genetics*, Vol. 12, p. e1006332. doi: 10.1371/journal.pgen.1006332

Mahmood, K., Mathiassen, S. K., Kristensen, M., & Kudsk, P. (2016). Multiple Herbicide Resistance in Lolium multiflorum and Identification of Conserved Regulatory Elements of Herbicide Resistance Genes. *Frontiers in Plant Science*, *7*, 1160.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., … Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, *461*(7265), 747–753.

Marouli, E., Graff, M., Medina-Gomez, C., Lo, K. S., Wood, A. R., Kjaer, T. R., … Lettre, G. (2017). Rare and low-frequency coding variants alter human adult height. *Nature*, *542*(7640), 186–190.

Matthew J. Foes, Tranel, P. J., Loyd M. Wax, & Edward W. Stoller. (1998). A Biotype of Common Waterhemp (Amaranthus rudis) Resistant to Triazine and ALS Herbicides. *Weed Science*, *46*(5), 514–520.

McVean, G., & Auton, A. (2007). LDhat 2.1: a package for the population genetic analysis of recombination. *Department of Statistics, Oxford, OX1 3TG, UK*. Retrieved from https://pdfs.semanticscholar.org/37c0/71c2bcd115bcce4bb2e4eb89a7dbb861c136.pdf

Mishra, P., & Panigrahi, K. C. (2015). GIGANTEA â€“ an emerging story. *Frontiers in Plant Science*, Vol. 6. doi: 10.3389/fpls.2015.00008

Molin, W. T., Wright, A. A., Lawton-Rauh, A., & Saski, C. A. (2017). The unique genomic landscape surrounding the EPSPS gene in glyphosate resistant Amaranthus palmeri: a repetitive path to resistance. *BMC Genomics*, *18*(1), 91.

Moretti, M. L., Van Horn, C. R., Robertson, R., Segobye, K., Weller, S. C., Young, B. G., … Others. (2018). Glyphosate resistance in Ambrosia trifida: Part 2. Rapid response physiology and non-target-site resistance. *Pest Management Science*, *74*(5), 1079–1088.

Nakka, S., Godar, A. S., Wani, P. S., Thompson, C. R., Peterson, D. E., Roelofs, J., & Jugulam, M. (2017). Physiological and Molecular Characterization of Hydroxyphenylpyruvate Dioxygenase (HPPD)-inhibitor Resistance in Palmer Amaranth (Amaranthus palmeri S.Wats.). *Frontiers in Plant Science*, *8*, 555.

Nandula, V. K., Ray, J. D., Ribeiro, D. N., Pan, Z., & Reddy, K. N. (2013). Glyphosate Resistance in Tall Waterhemp (Amaranthus tuberculatus) from Mississippi is due to both Altered Target-Site and Nontarget-Site Mechanisms. *Weed Science*, *61*(3), 374–383.

Neve, P., Vila-Aiub, M., & Roux, F. (2009). Evolutionary-thinking in agricultural weed management. *The New Phytologist*, *184*(4), 783–793.

Novembre, J., & Barton, N. H. (2018). Tread Lightly Interpreting Polygenic Tests of Selection. *Genetics*, *208*(4), 1351–1355.

Pan, L., Yu, Q., Han, H., Mao, L., Nyporko, A., Fan, L., … Powles, S. (2019). Aldo-keto Reductase Metabolizes Glyphosate and Confers Glyphosate Resistance in Echinochloa colona. *Plant Physiology*, *181*(4), 1519–1534.

Patzoldt, W. L., & Tranel, P. J. (2007). Multiple ALS Mutations Confer Herbicide Resistance in Waterhemp (Amaranthus tuberculatus). *Weed Science*, *55*(5), 421–428.

Patzoldt, W. L., Tranel, P. J., & Hager, A. G. (2002). Variable herbicide responses among Illinois waterhemp (Amaranthus rudis and A. tuberculatus) populations. *Crop Protection* ,

21(9), 707–712.

Patzoldt, W. L., Tranel, P. J., & Hager, A. G. (2005). A waterhemp (Amaranthus tuberculatus) biotype with multiple resistance across three herbicide sites of action. *Weed Science*, *53*(1), 30–36.

Peng, Y., Abercrombie, L. L., Yuan, J. S., Riggins, C. W., Sammons, R. D., Tranel, P. J., & Stewart, C. N., Jr. (2010). Characterization of the horseweed (Conyza canadensis) transcriptome using GS-FLX 454 pyrosequencing and its application for expression analysis of candidate non-target herbicide resistance genes. *Pest Management Science*, *66*(10), 1053–1062.

Preston, C. (2004). Herbicide resistance in weeds endowed by enhanced detoxification: complications for management. *Weed Science*, *52*(3), 448–453.

Preston, C., Tardif, F. J., Christopher, J. T., & Powles, S. B. (1996). Multiple Resistance to Dissimilar Herbicide Chemistries in a Biotype ofLolium rigidumDue to Enhanced Activity of Several Herbicide Degrading Enzymes. *Pesticide Biochemistry and Physiology*, *54*(2), 123–134.

Racimo, F., Berg, J. J., & Pickrell, J. K. (2018). Detecting Polygenic Adaptation in Admixture Graphs. *Genetics*, *208*(4), 1565–1584.

Riggins, C. W., Peng, Y., Stewart, C. N., Jr, & Tranel, P. J. (2010). Characterization of de novo transcriptome for waterhemp (Amaranthus tuberculatus) using GS-FLX 454 pyrosequencing and its application for studies of herbicide target-site genes. *Pest Management Science*, *66*(10), 1042–1052.

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., Schaffner, S. F., … Lander, E. S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, *419*(6909), 832–837.

Sabina, R. L., Paul, A.-L., Ferl, R. J., Laber, B., & Lindell, S. D. (2007). Adenine nucleotide pool perturbation is a metabolic trigger for AMP deaminase inhibitor-based herbicide toxicity. *Plant Physiology*, *143*(4), 1752–1760.

Salas-Perez, R. A., Saski, C. A., Noorai, R. E., Srivastava, S. K., Lawton-Rauh, A. L., Nichols, R. L., & Roma-Burgos, N. (2018). RNA-Seq transcriptome analysis of Amaranthus palmeri with differential tolerance to glufosinate herbicide. *PloS One*, *13*(4), e0195488.

Serra, A.-A., Nuttens, A., Larvor, V., Renault, D., Couée, I., Sulmon, C., & Gouesbet, G. (2013). Low environmentally relevant levels of bioactive xenobiotics and associated degradation products cause cryptic perturbations of metabolism and molecular stress responses in Arabidopsis thaliana. *Journal of Experimental Botany*, Vol. 64, pp. 2753–2766. doi: 10.1093/jxb/ert119

Shergill, L. S., Bish, M. D., Jugulam, M., & Bradley, K. W. (2018). Molecular and Physiological Characterization of Six-way Resistance in an Amaranthus tuberculatus var. rudis biotype from Missouri. *Pest Management Science*. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/ps.5082

Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redoña, E., … McCouch, S. R. (2015). Genomic selection and association mapping in rice (Oryza sativa): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genetics*, *11*(2), e1004982.

Swarts, K., Gutaker, R. M., Benz, B., Blake, M., Bukowski, R., Holland, J., … Burbano, H. A. (2017). Genomic estimation of complex traits reveals ancient maize adaptation to temperate

North America. *Science*, *357*(6350), 512–515.

Szpiech, Z. A., & Hernandez, R. D. (2014). selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Molecular Biology and Evolution*, *31*(10), 2824–2827.

Tennessen, J. A., Bigham, A. W., O'Connor, T. D., Fu, W., Kenny, E. E., Gravel, S., … NHLBI Exome Sequencing Project. (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*, *337*(6090), 64–69.

Tranel, P. J., Riggins, C. W., Bell, M. S., & Hager, A. G. (2011). Herbicide resistances in Amaranthus tuberculatus: a call for new options. *Journal of Agricultural and Food Chemistry*, *59*(11), 5808–5812.

Van Etten, M., Lee, K. M., Chang, S.-M., & Baucom, R. S. (2019). Parallel and nonparallel genomic responses contribute to herbicide resistance in Ipomoea purpurea, a common agricultural weed (p. 647164). doi: 10.1101/647164

Van Horn, C. R., Moretti, M. L., Robertson, R. R., Segobye, K., Weller, S. C., Young, B. G., … Gaines, T. A. (2018). Glyphosate resistance in Ambrosia trifida: Part 1. Novel rapid cell death response to glyphosate. *Pest Management Science*, *74*(5), 1071–1078.

Varanasi, V. K., Brabham, C., & Norsworthy, J. K. (2018). Confirmation and Characterization of Non–target site Resistance to Fomesafen in Palmer amaranth (Amaranthus palmeri). *Weed Science*, *66*(6), 702–709.

Vencill, W. K., Grey, T. L., Culpepper, A. S., Gaines, C., & Westra, R. (2008). Herbicide-resistance in the Amaranthaceae. *Journal of Plant Diseases and Protection: Scientific Journal of the German Phytomedical Society* , 41–44.

Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biology*, *4*(3), e72.

Watanabe, K., Stringer, S., Frei, O., Umićević Mirkov, M., de Leeuw, C., Polderman, T. J. C., … Posthuma, D. (2019). A global overview of pleiotropy and genetic architecture in complex traits. *Nature Genetics*, *51*(9), 1339–1348.

Weir, B. S., & Cockerham, C. C. (1984). ESTIMATING F-STATISTICS FOR THE ANALYSIS OF POPULATION STRUCTURE. *Evolution; International Journal of Organic Evolution*, *38*(6), 1358–1370.

Yang, Z., Gong, Q., Wang, L., Jin, Y., Xi, J., Li, Z., … Li, F. (2018). Genome-Wide Study of YABBY Genes in Upland Cotton and Their Expression Patterns under Different Stresses. *Frontiers in Genetics*, Vol. 9. doi: 10.3389/fgene.2018.00033

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

Yu, Q., Abdallah, I., Han, H., Owen, M., & Powles, S. (2009). Distinct non-target site mechanisms endow resistance to glyphosate, ACCase and ALS-inhibiting herbicides in multiple herbicide-resistant Lolium rigidum. *Planta*, *230*(4), 713–723.

Yu, Q., Cairns, A., & Powles, S. (2007). Glyphosate, paraquat and ACCase multiple herbicide resistance evolved in a Lolium rigidum biotype. *Planta*, *225*(2), 499–513.

Zhao, B., Huo, J., Liu, N., Zhang, J., & Dong, J. (2018). Transketolase Is Identified as a Target of Herbicidal Substance α-Terthienyl by Proteomics. *Toxins*, *10*(1). doi: 10.3390/toxins10010041

Zhao, N., Yan, Y., Ge, L. 'an, Zhu, B., Liu, W., & Wang, J. (2019). Target site mutations and cytochrome P450s confer resistance to fenoxaprop-P-ethyl and mesosulfuron-methyl in Alopecurus aequalis. *Pest Management Science*, *75*(1), 204–214.

Zhou, X., Carbonetto, P., & Stephens, M. (2013). Polygenic modeling with bayesian sparse linear mixed models. *PLoS Genetics*, *9*(2), e1003264.

Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, *44*(7), 821–824.

Supplemental Tables

Sup Table 1. List of individuals within populations and geographic regions used in the GWAS, along with average glyphosate resistance ratings, counts of TSR, and mean EPSPS copy number.

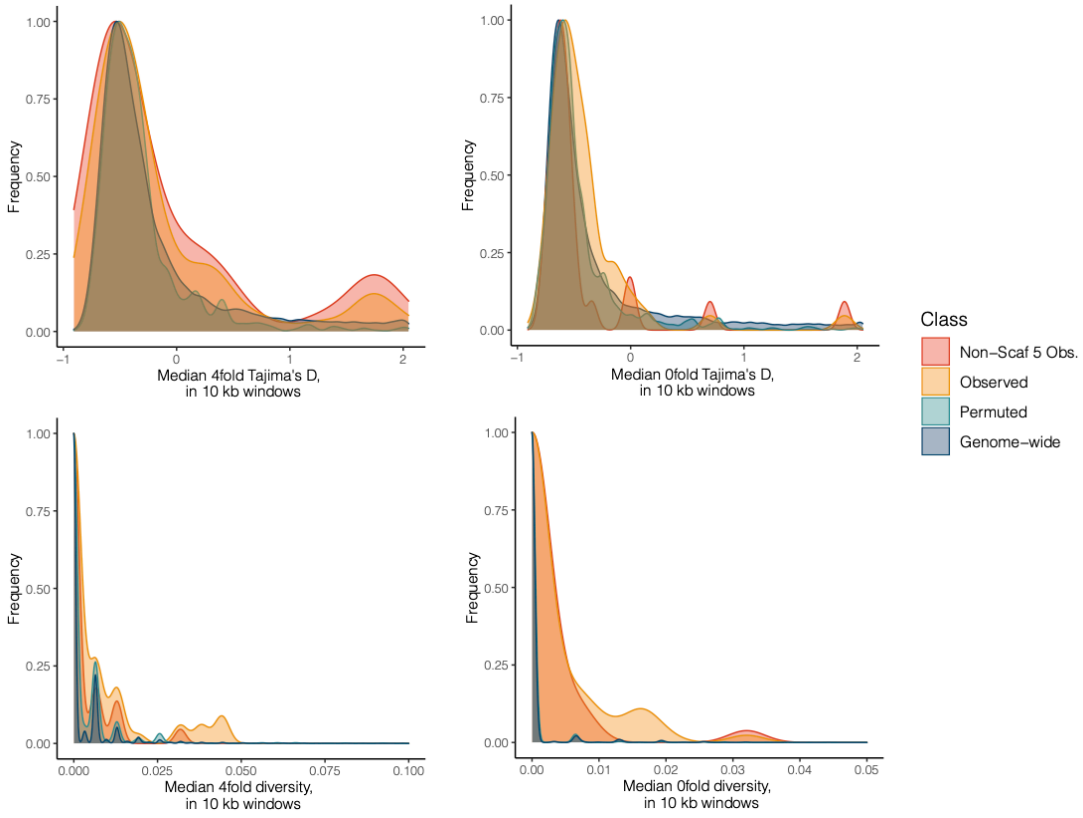
Region	Population ID	n	Mean Resistance Rating (1-5)	n (TSR)	Mean EPSPS Copy Number
Walpole	15	6	1.33	0	1.19
	16	4	1.00	0	1.18
	17	8	3.00	0	9.63
	18	6	3.33	0	10.36
	19	3	2.67	1	6.72
	27	8	2.38	0	3.85
Essex	4	10	1.40	0	1.60
	5	9	2.44	0	4.64
	6	6	1.67	1	1.40
	7	6	2.50	1	5.44
	9	7	1.14	0	1.09
Illinois	B	10	2.60	0	1.45
	D	10	2.30	6	1.11
	E	8	3.00	0	4.61
	F	10	2.30	0	4.59
	G	7	3.71	1	5.07
	H	10	2.40	0	3.76
	J	10	2.30	0	5.31
Missouri	K	9	2.22	0	3.88
Natural	N	8	1.00	0	1.12

Sup Table 2. Panther GO Enrichment results for Biological Processes, Protein Class, and Cellular Component for our significant SNP set from the TSR model (196 SNPs, 125 *Amaranthus tuberculatus* genes, 70 *Arabidopsis* orthologues).

PANTHER GO-Slim Cellular Component	Total	Observed	Expected	Fold Enrichment	raw P value	FDR
endocytic vesicle	2	1	0.01	> 100	5.09E-03	1
lipid droplet	5	1	0.01	78.33	1.27E-02	1
autophagosome	5	1	0.01	78.33	1.27E-02	1
vacuolar membrane	127	2	0.32	6.17	4.21E-02	1
vacuolar part	127	2	0.32	6.17	4.21E-02	1
endosome	134	2	0.34	5.85	4.64E-02	1
PANTHER GO-Slim Biological Process						
acylglycerol catabolic process	1	1	0	> 100	5.09E-03	1
glycerolipid catabolic process	4	1	0.01	97.91	1.27E-02	1
acylglycerol metabolic process	13	1	0.03	30.13	3.51E-02	1
neutral lipid metabolic process	13	1	0.03	30.13	3.51E-02	1
neutral lipid catabolic process	1	1	0	> 100	5.09E-03	1
mitochondrial transcription	3	1	0.01	> 100	1.01E-02	1
mitochondrial RNA metabolic process	19	1	0.05	20.61	4.97E-02	1
cell fate specification	5	1	0.01	78.33	1.52E-02	1
cell fate commitment	19	1	0.05	20.61	4.97E-02	1
IMP metabolic process	6	1	0.02	65.28	1.77E-02	1
nucleoside monophosphate metabolic process	14	1	0.04	27.98	3.75E-02	1
regulation of circadian rhythm	8	1	0.02	48.96	2.27E-02	1
regionalization	9	1	0.02	43.52	2.52E-02	1

pattern specification process	9	1	0.02	43.52	2.52E-02	1
circadian rhythm	12	1	0.03	32.64	3.26E-02	1
rhythmic process	12	1	0.03	32.64	3.26E-02	1
ammonium ion metabolic process	12	1	0.03	32.64	3.26E-02	1
triglyceride metabolic process	13	1	0.03	30.13	3.51E-02	1
fruit development	14	1	0.04	27.98	3.75E-02	1
lipid homeostasis	31	2	0.08	25.27	3.21E-03	1
glucose 6-phosphate metabolic process	16	1	0.04	24.48	4.24E-02	1
leaf development	18	1	0.05	21.76	4.73E-02	1
PANTHER Go-Slim Protein Class						
transketolase	2	1	0	> 100	4.26E-03	2.43E-01
transferase	682	5	0.97	5.15	2.75E-03	2.35E-01
metabolite interconversion enzyme	2275	10	3.24	3.09	1.05E-03	1.80E-01
deaminase	16	1	0.02	43.94	2.39E-02	6.81E-01
hydrolase	492	3	0.7	4.29	3.30E-02	8.05E-01
exoribonuclease	25	1	0.04	28.12	3.63E-02	7.76E-01
decarboxylase	26	1	0.04	27.04	3.77E-02	7.16E-01

Supplementary Figures



Sup Figure 1. Distributions of 4-fold and 0-fold Tajima's D and diversity in 10kb windows surrounding focal resistance-related SNPs (across all scaffolds, or excluding the EPSPS-containing scaffold 5) relative to the genome-wide and permuted distributions.