The genetic architecture and genomic context of glyphosate resistance in

Amaranthus tuberculatus

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*

[†]these authors contributed equally

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

Molecular Ecology Special Issue

February 2021

Abstract

Much of what we know about the genetic basis of herbicide resistance has come from detailed investigations of monogenic adaptation at known target-sites, despite the increasingly recognized importance of polygenic resistance. Little work has been done to characterize the broader genomic basis of herbicide resistance, including the number and distribution of genes involved, their effect sizes, allele frequencies, and signatures of selection. In this work, we implemented genome-wide association (GWA) and population genomic approaches to examine the genetic architecture of glyphosate resistance in the problematic agricultural weed, *Amaranthus tuberculatus*. A GWA was able to correctly identify the gene targeted by glyphosate, but when we statistically controlled for two target-site genetic mechanisms, we found an additional 250 genes across all 16 chromosomes associated with non-target site resistance (NTSR). The encoded proteins had functions that have been linked to non-target site resistance (NTSR), the most significant of which is response to chemicals, but also showed pleiotropic roles in reproduction and growth. The architecture of NTSR was enriched for large effect sizes and low allele frequencies, suggesting the role of pleiotropic constraints on its evolution. The enrichment of rare alleles also suggested that the genetic architecture of NTSR may be population-specific and heterogeneous across the range. Despite their rarity, we found signals of recent positive selection on NTSR-alleles by both window- and haplotype-based statistics, and an enrichment of amino-acid changing variants. In our samples, genome-wide SNPs explain a comparable amount of the total variation in glyphosate resistance to monogenic mechanisms, even in a collection of individuals where 80% of resistant individuals have large-effect TSR mutations, indicating 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 gene products. In practice, however, herbicide resistance in evolved weed populations is complex. Agricultural weed populations evolve resistance 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 agricultural weed populations (Heap, 2010), we are only beginning to learn about the number of loci involved and the associated distribution of allelic effect sizes, frequencies, and the fraction of phenotypic variance explained by NTSR (Van Etten, Lee, Chang, & Baucom, 2019). 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, uncovering the genetics of NTSR will contribute to our understanding of the relative importance and consistency of small- and large-effect alleles to rapid adaptation, while also shedding light on how strong selection shapes genome-wide variation of natural weed populations.

NTSR can be broadly defined as all alleles that provide tolerance or resistance to herbicides unexplained by and unlinked to TSR, regardless of mechanism—by providing direct protective effects, compensentating for herbicide toxicity through downstream pathways, or mediating the costs and/or benefits of TSR alleles—and we use this conceptual approach in this work. It is likely that such mechanisms are similarly important as TSR in conferring herbicide resistance across agriculturally important ranges (e.g. Delye, Gardin, Boucansaud, Chauvel, & Petit, 2011),

as several studies have looked for causative mutations in the gene encoding the herbicide-targeted protein in resistant populations, but failed to find one (Délye, 2013; Guo et al., 2015; Van Etten et al., 2019; Van Horn et al., 2018). In contrast to a single large-effect TSR locus,

65 where resistance alleles are often assumed to have arisen *de novo*, the polygenic basis of resistance with many small-effect loci is more likely to draw from genome-wide standing genetic variation (Kreiner et al., 2018; Paul 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

70 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

75 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

80 (Yuan, Tranel, & Stewart, 2007). These NTSR genes are thought to act by altering herbicide penetration, translocation, accumulation at the target site, as well as through offering protection from herbicide effects (Délye, 2013; Moretti et al., 2018), the function of several of which have been experimentally validated (Ito & Gray, 2006; Xi, Xu, & Xiang, 2012). However, the genomic, rather than transcriptomic, basis of NTSR is unknown in most species and for most

85 classes of herbicides (but see 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 et al., 2018; Leslie & Baucom, 2014; Molin, Wright, Lawton-Rauh, & Saski, 2017; Van Etten et al., 2019), traditional

90 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

2

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

95   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^{TM}, are widely

100   used across North America to suppress weed populations, especially in soy and corn fields. Glyphosate targets a step in the synthesis of aromatic amino acids, carried out by 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS). These aromatic amino acids are required not just for protein synthesis, but for the synthesis of many natural compounds such as hormones, pigments, and cell wall components, and their production draws on more than 30% of

105   photosynthetically fixed carbon (Maeda & Dudareva, 2012). Glyphosate herbicides were 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 has increased exponentially, and is third to only ALS-inhibiting and Photosystem II-inhibiting herbicides in the number of resistant species (Heap, 2010). Meanwhile,

110   glyphosate is still the most widely used herbicide across North America—due to 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).

115   *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

120   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

3

at EPSPS and its related amplification in response to glyphosate, across natural and agricultural

125   populations (Kreiner et al., 2019). Our previous work was a targeted investigation of repeated

origins of monogenic target-site resistance, particularly signals of positive selection from

glyphosates around the EPSPS gene in individuals with high EPSPS copy number but did not

attempt to characterize polygenic signals of selection for herbicide resistance. With TSR

mechanisms incompletely explaining phenotypic resistance levels in those samples (Kreiner et

130   al., 2019), and reports of non-target site resistance to glyphosate in the species (Nandula, Ray,

Ribeiro, Pan, & Reddy, 2013), in the current study we sought to identify genome-wide effects of

resistance and signals of adaptation to herbicides.

We implemented GWA and population genomic approaches for characterizing the genetic

135   architecture of glyphosate resistance in *Amaranthus tuberculatus*. Specifically, we aimed not

only to identify putative non-target site loci across the genome, but also to explore the relative

importance of this polygenic architecture compared to previously characterized target-site

resistance mutations in our dataset. Furthermore, we sought to understand modes of selection on

polygenic resistance. As expected, a GWA for resistance to glyphosate in *Amaranthus*

140   *tuberculatus* without TSR covariates shows the largest genome-wide peak on chromosome 5

near the known-target, EPSPS. To identify putative NTSR loci, we included both of the known

TSR mechanisms as covariates in the GWAS (as in Atwell et al., 2010; Sasaki, Köcher, Filiault,

& Nordborg, n.d.); this revealed hundreds of other alleles with genome-wide significance after

correcting for multiple testing (FDR) and relatedness among individuals. Furthermore, we

145   compared our observations to an empirical null that allowed us to evaluate the extent to which

differences in the genetic architecture of resistance across populations may be confounded with

population structure. Against this null expectation, we asked whether the architecture we

identified was enriched for relevant NTSR-related functions and population genomic signals of

selection, and evaluated its importance against widespread TSR mechanisms. This work

150   contributes to our understanding of the heterogeneity in the genetic architecture of resistance

across populations, its importance compared to TSR mechanisms, and selective processes that

have governed its evolution.

Materials & Methods

155

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)

160   (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 genomic

samples, collected from 8 fields with high *A. tuberculatus* densities across Missouri and Illinois

165   in the Midwest United States, and from counties with newly resistant populations 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

170   3-5 offspring individuals per line of field-collected specimens. 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 percent injury, with 1 indicating

complete plant necrosis and 5 indicating no apparent damage. A subset of these phenotyped

offspring individuals were then selected for resequencing, with the goal of representing both

175   susceptible and resistant individuals across populations (i.e., the distribution of resistance within

our set of sequenced individuals was not chosen to reflect population frequencies). Individual-

level resistance ratings were directly used in the GWAS, as opposed to family-level means, as

variance in resistance within a family may be encompassed by differences at many loci genome-

wide. We analyzed GWAs with resistance using these discrete 1-5 ratings.

180

Genetic TSR characterization & Multiple Regression

We previously characterized TSR mechanisms in (Kreiner et al., 2019). Briefly, after SNP

calling and filtering, we located the known TSR mutation codon *EPSPS* Pro106Ser by

submitting a validated EPSPS complete mRNA coding sequence (genbank sequence:

5

185   FJ869881.1) to a BLAST (Altschul, Gish, Miller, Myers, & Lipman, 1990) search to our reference genome. Individuals containing the Pro106Ser mutation were all heterozygous, and thus the *EPSPS* Pro106Ser covariate was coded based on presence/absence (0/1). Individual EPSPS copy number was estimated by dividing the median coverage at every base-pair position within the EPSPS gene, by the median genome-wide coverage (since the median coverage

190   should be a good proxy for coverage of single copy genes genome-wide) and taking the gene-wide average.

For estimating the proportion of phenotypic variance in resistance explained (PVE) by TSR mechanisms, we used a simple linear regression of resistance on Pro106Ser presence/absence

195   and *EPSPS* copy number. We decided not to include the kinship matrix as a covariate in this regression, as validated large-effect TSR mutations are unlikely to be driven by population structure and our aim was to avoid attributing variance in resistance explained by TSR to genome-wide relatedness. This should lead to a relatively conservative estimate of the importance of NTSR mechanisms compared to TSR mechanisms. While ideally, we would be

200   able to extract this information directly from the GWA of NTSR that includes both genetic covariates, it is infeasible to do so given current GWA software packages, which often have the goal of controlling for covariates rather than estimating the variance explained as an object of study.

205   To estimate the SNP heritability of glyphosate resistance, we additionally used the R package "lme4qtl" to run a multiple regression of resistance onto both TSR genetic mechanisms including the kinship matrix used in our GWA as a random-effect covariate, using the function "relmatLmer".

210   *Linear mixed GWA models & preprocessing & GO enrichment*
Filtered VCFs were obtained from (Kreiner et al., 2019). Briefly, freebayes-called SNPs were filtered based on missing data (<=20% missing), 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

215   these 162 samples showed some individuals with particularly high levels of relatedness, some of

6

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.

We implemented two linear-mixed-model GWA tests, first to assess whether GWA could
220 identify the large effect TSR locus (EPSPS), and the next to discover NTSR-related alleles genome-wide. The first model had only raw resistance ratings as the response variable, with no covariates. The second model included two genetic covariates: a TSR mutation of proline to serine at codon number 106 in *EPSPS* (P106S), and the magnitude of the copy number of *EPSPS* (as characterized in Kreiner et al., 2019). We included these genetic covariates to ensure that
225 variance in resistance explained by TSR mechanisms was not attributed to other alleles across the genome and that differences in the genetic architecture of resistance across populations does not lead to spurious associations with population structure (as in Atwell et al., 2010; Sasaki, Köcher, Filiault, & Nordborg, n.d.). We did so in GEMMA (Zhou & Stephens, 2012), using the lmm -4 option and after estimation of a kinship/relatedness matrix (-gk), with covariates
230 specified (-c option) when relevant. We employed both a Bonferroni correction and an FDR approach to significance testing in R through the function p.adjust (method="bonferroni" or "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
235 be overly conservative in a GWA setting due to the assumption of independence between tests (Johnson et al., 2010), especially true in our case as we chose to not implement LD thinning.

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 GO molecular
240 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
245 from several species to identify known genes in 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.

250   *GWA permutations*

To generate a null distribution of expected significant hits, and as a point of comparison for downstream population genetic analyses for the two covariate NTSR model, we randomized phenotype with respect to multilocus genotypes. We randomized the phenotype of interest and covariates together with respect to SNP multilocus genotypes, such that the associations between

255   the resistance and covariates were maintained—this provides an expectation for ascertainment bias of significant SNPs in regions of low recombination as well as avoids circularity between GWA ascertainment and downstream population genomic analyses. Furthermore, to incorporate the effects of population structure that may have gone uncontrolled by the kinship matrix we performed this joint randomization in two increasingly stringent ways; within geographic regions

260   (Ontario Natural Populations, Essex, Walpole, Illinois, Missouri; where the main axis of population structure has been shown to differ; Kreiner et al., 2019), and more rigorously, within individual populations. Therefore, our empirical null distribution represents the expected results from a GWA with our given phenotypic distribution among populations, stratification between populations, and after controlling for individual relatedness.

265

For downstream analyses of evidence of selection, while we present distributions of significant SNPs for both permutations, our tests of significance are performed against the more rigorous within-population permutations. To do so, for each permutation type, we generated 250 sets of randomized phenotype/covariate to genotype data, running a GWA on each. To generate null

270   distributions of the expected number of significant SNPs, we used the significance threshold (q value < 0.05) observed in the two covariate GWA model. Based on the number of significant SNPs extracted from each randomized iteration, we then calculated the average number of significant SNPs across all 250 permutations to get an 'empirical' FDR. For illustrative purposes we plotted the distribution of various statistics for all significant SNPs across permuted GWAs

275   for both the within-region and within-population permutations (i.e., in **Fig 3**), and so it is important to note that these illustrated distributions reflect very different sample sizes (~500 for the observed set and ~60,000 for the permuted set). For rigorous significance testing, we relied

on the more stringent within-population permutations, resampling the total pool of significant

SNPs across all permuted GWAs to the same number of observations we observed in our true

280   GWAS, 1000 times. For each resample, we took the median (or mean) of a given statistic, 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 GWA. We also investigated the

extent to which significant SNPs from each within-population permutation was enriched for

285   biological classes after bonferroni correction, as described above for our observed GWA.

Summary statistics

To infer patterns of selection on putative NTSR loci and characteristics of the genetic

architecture of resistance, we estimated a variety of population genomic statistics on significant

290   hits. Allele frequency and effect sizes were analyzed directly from the output of the GEMMA

GWA. Since linked selection should reduce diversity at 4-fold degenerate coding sites

(putatively neutral) relative to 0-fold degenerate coding sites (putatively selective) (Smith &

Haigh, 1974), we estimated diversity at these sites independently in 10 kb genomic windows

around focal alleles. We used scripts from https://github.com/simonhmartin/genomics_general to

295   centre windows on putative NTSR and permuted significant (false positive) SNPs, including

both variant and invariant 0-fold or 4-fold sites. For both visualization and statistical resampling,

we required windows to have data for at least 0.01% of the window (10/10000 invariant or

variant sites). Tajima's D (a summary of the extent of neutral evolution) was calculated at 0fold

and 4fold sites and was estimated on a per-SNP basis, and median genomic windows were

300   estimated with custom dplyr functions.

We also calculated three related haplotype-based statistics—extended haplotype homozygosity

(EHH; Sabeti et al., 2002), integrated haplotype homozygosity (iHH), and the integrated

haplotype score—that can be more powerful than single SNP or windowed approaches for

305   identifying positive selection as they draw power from signals of selection at linked sites. All of

such statistics are 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

9

had a negative effect on resistance, such that all 1 alleles correspond to alleles with positive

310   effects on resistance. These 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 the iHH (integrating the space under the EHH curve) by

taking the difference in iHH among haplotypes and standardizing for allele frequency in bins

315   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.

320

All custom scripts are available at https://github.com/jkreinz/NTSR-GWAS-MolEcol2020.

Results

SNPs and Traits

325   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 the EPSPS

Pro106Ser substitution (Fig 1). Out of all 155 individuals, 81 were classified as glyphosate

330   resistant (a rating of ≥ 2 on a scale of 1 to 5, i.e., less than ~80% plant injury) and 74 as

susceptible. Note that the 1-to-5 scale was used as input for the GWA tests. Only ten out of 155

individuals had a TSR mutation in the *EPSPS* coding sequence, while the *EPSPS* amplification

was much more prevalent: 79 individuals had a copy number ≤ 1.5, while the copy number in 76

individuals exceeded 1.5 (medians of 1.13 and 5.95 copies, respectively), with one individual

335   ranging up to 30 copies of *EPSPS*. Seventeen individuals have no evidence for either TSR

mechanism but were rated as resistant (Fig 1A**;** "Res., noTSR"), resistance being most likely

driven solely by NTSR mechanisms. Moreover, a regression of resistance on both TSR

mechanisms (EPSPS Pro106Ser and *EPSPS* copy number) explained only 33% of the variance in

resistance ratings ($r^2 = 0.33$, $F_{2,152} = 38.69$, $p = 2.62e-14$), implying that NTSR mechanisms (Fig

10

340   1B "NTSR") may be widespread even across individuals with at least one TSR mechanism,

which were found in 79% (64/81) of our resistant samples. Thus, genome-wide loci may be

contributing both to resistance in the absence of TSR, and to the degree of resistance conferred

by individuals with a TSR mechanism. In fact, because some samples with TSR were not

phenotyped as resistant, coexistence of NTSR mechanisms may be required for robust
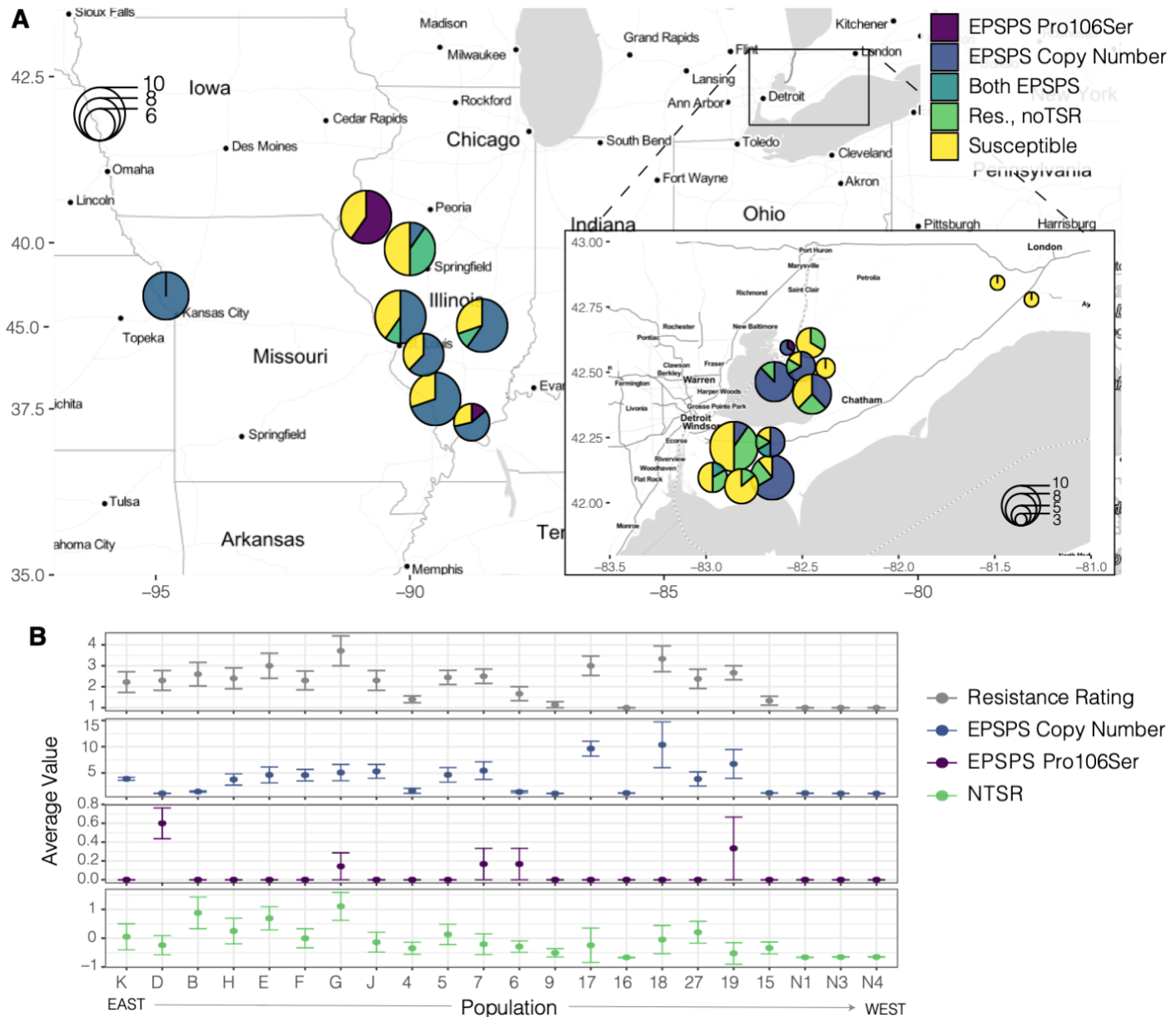
345   phenotypic resistance.

Figure 1. Distribution of glyphosate resistance types across populations. A) Map of populations included in the GWA, showing the proportion of susceptible and resistant individuals of different underlying genetic mechanisms based on presence/absence. "Res., noTSR" refers to individuals rated as resistant at least 2 on a 1-to-5 scale but that lack evidence for one of the two known TSR mechanisms. B) Population-means of resistant types. The NTSR rating represents the residuals from a regression of resistance on both TSR mechanisms (*EPSPS* copy number and *EPSPS* Pro106Ser), where the lowest values on the y-axis are completely susceptible. Error bars represent standard deviation of the population mean estimate.

380   *Gemma Linear Mixed Models & PVE*

We performed two GWA mixed effect models, both of which included a random effect kinship matrix but differed in the presence of any fixed effect, genetic covariates (*EPSPS* Pro106Ser and *EPSPS* Copy Number, Table 1). The no-covariate model was run to determine whether a GWA for our resistance ratings could accurately identify the large-effect target-site resistance gene,

385   *EPSPS*, while the two covariate model (NTSR model) was run to identify additional candidate regions across the genome contributing to polygenic resistance, controlling for any linkage effects with TSR and/or differences in the genetic architecture of resistance among populations that might lead to spurious associations between TSR and population structure. The no-covariate model had a lambda of 0.979 and the qq-plot showed a large tail of an excess observed-to-

390   expected p-values. While the elevated tail in the qq-plot was less pronounced for the NTSR two-covariate GWA model, the lambda estimate was the same, and this model was associated with a higher maximum likelihood estimate (Table 1, Sup Fig 1).

The no-covariate model had 91 significant SNPs after correction, compared to 574 SNPs in the

395   two covariate, NTSR model. The global-, regional-, and population-level permutations provided estimates of the 'empirical' FDR based on the average number of SNPs that are significant at this threshold across 250 phenotype-genotype randomized GWAs. Despite using a stringent FDR cutoff of $\alpha = 0.05$ for significance testing in both the covariate-free and two-covariate model, we found with these permutations relatively large numbers of significant SNPs (55, 165, and 229 out

400   of the 573, respectively) (Table 1). The number of significant SNPs in permutations suggested a residual effect of population structure unaccounted for by the initial significance cutoff and controlling for the kinship matrix—or put another way, that differences in the genetic architecture of resistance among populations covary with fine-scale population structure. While these empirical estimates of the FDR highlight that a number of these SNPs we characterize here

405   may result from either of these processes and individually should be treated with caution, we examine evidence for selection and relevant biological functions against the most stringent within-population permuted set, and provide a conservative estimate for the amount of variance in resistance that can be explained by these putatively NTSR-related alleles.

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

Covariates	-log10(p) at q = 0.05	# Sig. SNPs @ FDR	PVE† (SE)	MLE	Lambda (95% CI)	Global Permuted FDR estimate	Within-Region Permuted FDR estimate	Within-Pop Permuted FDR estimate
None	6.4	91	0.388 (0.230)	-268	0.979 (0.972, 0.986)	0.011 (1/91)	0.04 (4/91)	0.06 (5/91)
EPSPS TSR SNP + Copy Num.	5.6	573	0.374 (0.219)	-234	0.979 (0.971, 0.986)	0.096 (55/573)	0.29 (165/573)	0.40 (229/573)

† when covariates are included, PVE represents percent of *residual* variance explained

The relative importance of monogenic versus polygenic mechanisms of resistance can be explored by comparing the percent of variance explained (PVE) by genome-wide SNPs in the two covariate NTSR model (polygenic effects on resistance) to the r^2 of glyphosate resistance on TSR and EPSPS copy number (monogenic effects). Controlling for both TSR genetic mechanisms leads to a marginal PVE estimate of 0.374, implying that genome-wide SNPs can explain 37% after accounting for resistance associated with TSR at EPSPS. The initial multiple regression model of resistance on monogenic mechanisms (TSR SNP + EPSPS copy number) results in an $r^2 = 0.33$ ($F_{2,152} = 38.69$, $p = 2.62e\text{-}14$). Thus, a rough estimation of the residual variance unexplained by TSR mechanisms ($1 - 0.33 = 0.67$) suggests that the NTSR loci we identify here can explain about a quarter (0.37×0.67) of the variance in resistance—implying that even in a sample of resistant individuals that nearly all segregate for glyphosate TSR, NTSR mechanisms may be of near equal importance.

We also implemented a multiple regression of resistance ratings on both TSR genetic covariates, including the kinship matrix as a random effect covariate to estimate the SNP heritability of resistance, finding 83% of variance in resistance can be explained by additive genetic effects. This provides further support for the argument that a significant component of the variation in resistance that is unexplained by TSR is due to loci distributed throughout the genome.

430

Distribution of genome-wide associations, candidate genes, and GO enrichment

The qualitative pattern of GWAs with glyphosate resistance differed greatly between the no-covariate and NTSR, two covariate models (Fig. 2a,b). In the no-covariate model, a butte of

435   significant SNPs within the confines of the EPSPS-related amplification is apparent on scaffold 5 (Fig. 2b) with glyphosate resistance associated loci found directly within EPSPS. In contrast, the NTSR two-covariate model shows no signal of association at EPSPS, as is expected given our EPSPS-based covariates. The NTSR model shows significant hits distributed across all 16 chromosomes, with particularly high concentration of significant hits on scaffold 1, 15, and 2

440   (42%, 18%, 10% of genome-wide significant hits, respectively) **(Fig 2b,c)**. Importantly, a number of genome-wide significant hits found in the no-covariate model become non-significant in the NTSR model. We interpret this to reflect the fact that population-level differences in the genetic basis of resistance (**Fig 1**) can lead to spurious genome-wide associations due to covarying population structure. A potential confounding of population structure and the genetic

445   basis of resistance highlights the importance of the two-covariate model, as well as the regional and within-population permutations, for identifying variants associated with NTSR.
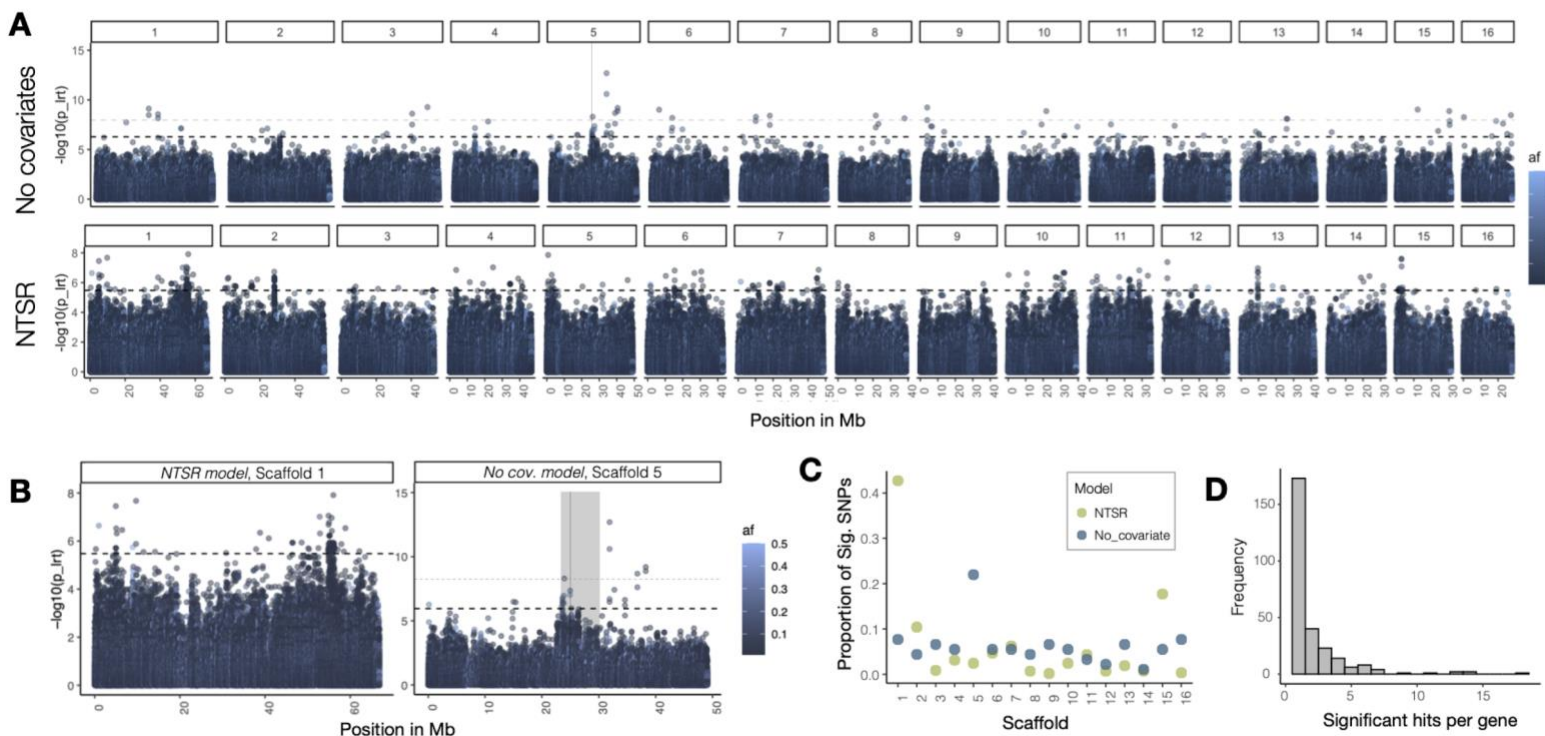
Figure. 2. GWA of glyphosate resistance. A) Manhattan plot of significance of association with glyphosate resistance across all 16 scaffolds for two linear mixed models - the upper without fixed effects and the lower with TSR as a covariate. Dashed black horizontal bar indicates the 0.05 FDR threshold, while the dashed grey bar above indicates the Bonferroni correction threshold. Thin vertical line on scaffold 5 depicts the EPSPS gene. X axis is scaled by the length of each scaffold. B) Manhattan plot for scaffold 1 (containing 42% of significant hits) [Left] as compared to the Manhattan plot for scaffold 5 in the no covariate model, which contains the target gene for glyphosate, EPSPS [Right]. 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 NTSR model.

The set of 573 significant SNPs with -log10(p) values corresponding to q < 0.05 (Table 1) in the NTSR model corresponded to 274 unique *A. tuberculatus* genes. Six genes have more than 10 significant SNPs mapping within them, four of which are found on Scaffold 1 in a peak between 54.9Mb and 56.3Mb (Fig 2b,c). These four genes are a DNA-directed DNA polymerase (AT5G44750), DNA-directed RNA polymerase family protein (AT2G15430), B-block binding subunit of TFIIIC (AT1G59077), and Auxin-responsive protein SAUR21 (AT5G18030) (Fig

16

465   2b). On Scaffold 2, a Vacuolar protein sorting-associated protein 26A (AT5G53530) also has >10 SNPs mapping.

For 150/274 of these *A. tuberculatus* genes, we could find mappable orthologs in *Arabidopsis thaliana*, and thus could be used as input for enrichment analyses and candidate gene
470   exploration. Fisher's exact test showed this set of genes to be significantly enriched after FDR correction for 13 distinct classes of GO biological processes in 6 hierarchical categories **(Fig 3)**, the two most significant independent terms being response to chemicals and biosynthetic processes.

475   Many of the biological processes of genes in our set of significant SNPs were notable given previous work on particular gene families and the three known molecular phases of NTSR (Gaines et al., 2020) (1 - herbicide detoxification via oxidation, hydrolysis, or reduction, 2 - conjugation, and 3 - compartmentalization/transport) and metabolic pathways known to be modulated by the Shikimate pathway targeted by glyphosate. The 573 alleles we found
480   significantly associated with NTSR function in relevant biological processes such as response to chemicals, stimulus, and response to oxygen-containing compounds, and numerous metabolism and biosynthetic and metabolic related processes (Fig 3). The highest proportion of NTSR-related genes function broadly in response to stimulus (55/150) or chemicals (38/150), and biosynthetic processes (35/150), all of which remain significant even after stringent bonferroni p-
485   value correction, along with 3 other carboxylic acid metabolic process-related classes (Fig 3). In comparison, across the 250 within-population GWA permutations, 225 have zero GO biological classes withstanding bonferroni correction, with the remaining iterations having only 1.17 terms passing bonferroni correction on average. Furthermore, none of these GWA iterations are significant for any of the GO biological processes we observe to pass bonferroni testing.
490

Key genes involved in response to chemicals and organic acid metabolic processes code for many protein classes, including aldolases (FBA6, FBA2 - regulate glycolytic processes in response to abiotic stress (Lu et al., 2012)), a dehydrogenase (CICDH - NADPH dependent dehydrogenase that plays a key role in redox homeostasis (Mhamdi et al., 2010)), decarboxylases
495   (ADC1), kinases (BSK2, BAM1, PI4KG4, ADK1, LIP1), reductases (GLO5, GRXC1 - involved

in redox regulation (Riondet et al., 2012)), methyltransferases (JMT - involved in plant defense, induced by wounding and methyl-jasmonate treatment (Seo et al., 2001)), WRKY transcription factors (WRKY27, WRKY1; known to play a key role in biotic and abiotic stress response (Pandey & Somssich, 2009; Phukan, Jeena, & Shukla, 2016)), auxin responsive genes (SAUR5;

500   SAUR50, DTX51; EXO70A1, ARF1), ethylene receptors (ETR1,ERF061), and an ABC transporter (ABCG39, part of the pleiotropic drug resistance (PDR) transporters, ABCG subfamily). A full list of Amaranthus and Arabidopsis genes corresponding to SNPs significantly correlated with NTSR can be found in supplement file 1.
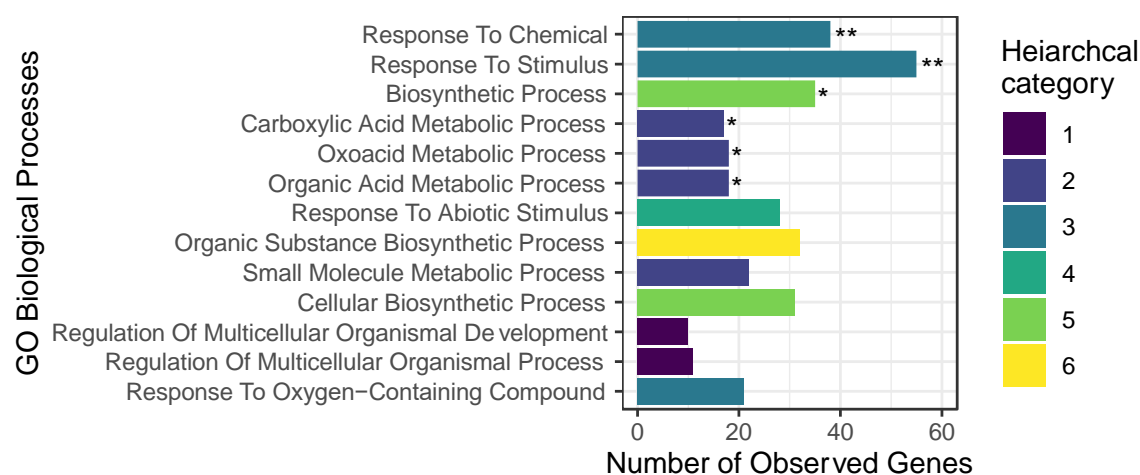

505   **Figure 3. 573 SNPs significantly associated with NTSR are enriched for 13 GO biological processes after FDR correction.** Classes are sorted by increasing raw p-value, and coloured by hierarchical categories. Categories significant after Bonferroni correction are indicated by asterisks (*** = p-value < 0.0001, ** < 0.01, * < 0.05).

510   While many of our NTSR-associated genes function in metabolic and chemical response pathways, we also see enrichment for categories less obviously related to detoxification— notably, regulation of multicellular organismal processes and development. Examples of the 11 genes enriched for these developmental processes include: *GIGANTEA*, encoding a nuclear protein with pleiotropic roles in flowering time, stress response, circadian clock regulation, with

515   prior evidence for providing herbicide tolerance by increasing oxidative stress resistance (Kurepa, Smalle, Va, Montagu, & Inzé, 1998; Mishra & Panigrahi, 2015); *EMBRYONIC FLOWER 2* (*EMF2*), which in *A. thaliana* prevents premature flowering (Yoshida et al., 2001), *PETAL LOSS* (*PTL*), which regulates perianth architecture (Brewer et al., 2004); *BBX18,* which suppresses photomorphogenesis by blue-light mediated hypocotyl reduction (Vaishak et al.,

18

520 2019; Wang et al., 2011), and *UNFERTILIZED EMBRYO SAC10* (*UNE10*) and *ARM REPEAT PROTEIN INTERACTING WITH ABF2* (*ARIA)*, two negative regulators of seed germination (Tepperman et al., 2004; Kim et al., 2004).

525 *Genomic context of putative NTSR alleles*

With a better understanding of the types of genes putatively involved in NTSR, we posited that a closer look at the genomic context of the loci in our set of candidate genes would shed light on the action of selection, genetic architecture, and heterogeneity of NTSR among populations. To

530 do so, we compared various characteristics of our observed NTSR-related SNPs to the within-population permuted expectation, such that particularly extreme values indicate evidence above and beyond the effects of genetic and phenotypic population structure or recombination heterogeneity across the genome.

535 Our reference genome was from a glyphosate-susceptible individual, and we polarized our SNPs against this reference. The effect sizes of non-reference alleles were highly asymmetric in direction; of the 573 non-reference variants, only 8 SNPs had negative effects on glyphosate resistance compared to 565 with positive effects. Our significant SNPs broadly fell into two categories, rare alleles of modest to large effect and common alleles with small effects (Fig. 3A).

540 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, Josephs, Lee, Stinchcombe, & Wright, 2015), a lack of power makes it difficult to detect rare, small-effect alleles, and both the Winner's Curse and

545 Beavis effect lead to an overestimation of detected effect sizes (Liu & Leal, 2012; Xu, 2003). The tradeoff between effect size and allele frequency was clearly present in the genetic architecture of NTSR, where our set of significant SNPs were enriched for large absolute effect sizes compared to the empirical null distribution (median beta = 2.455; [95% CIs of median empirical null distribution, based on resampling: 2.084, 2.186]), and for rare alleles (median AF

550 = 0.016, [95% CIs of median empirical null: 0.019, 0.023]) (Fig. 4A,B).

We tested for signals of directional selection on NTSR-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. When we truncated the EHH distribution to 10 kb around either side of observed resistance-associated alleles, we saw negligible differences in average EHH value across this region in observed resistance-, compared to permuted resistance-associated haplotypes (median R allele EHH across 20kb = 0.000473, 95% CIs of mean empirical null: 0.000, 0.0196) (Fig 3D). Taking the integral of the EHH by distance curve referred to as integrated Haplotype Homozygosity (iHH; Voight, Kudaravalli, Wen, & Pritchard, 2006)—but without truncating haplotype length—revealed that observed resistant haplotypes were longer than expected under the empirical permuted null distribution (median iHH = 0.042531, [95% CIs of median pop-permuted empirical null: 0.01552, 0.01929]) (Fig 3E). 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) identified enrichment for unusually large homozygous tracts of resistant relative to susceptible haplotypes, given their respective allele frequencies (median standardized iHS = 0.21286, [95% CIs of median pop-permuted empirical null: 0.00785, 0.13600]) (Fig 3F).

These haplotype-based inferences suggest a role for recent directional selection, and so as a final check we also estimated diversity at Tajima's D at 0-fold and 4-fold degenerate coding sites in 10 kb windows surrounding focal SNPs, where signatures of linked selection should show particularly strong signals of selection at putatively neutral 4-fold sites. We see no evidence for reduced 4-fold diversity or 0-fold diversity around NTSR-related SNPs compared to the permuted expectation (median 4fold π = 0.0439, [95% CIs: 0.0415, 0.0447]; median 0fold π = 0.0152759 [95% CIs: 0.01528749, 0.01728545]). However, whereas the observed SNP set had sufficient data (at least 10 variant or invariant degenerate sites of a given type) in 419/573 4-fold and 443/573 0-fold windows (therefore representing ~75% of relevant windows) the population permuted set had only 784/78036 and 814/78036 (~1%), limiting the inferences that can be drawn from this comparison. We see more negative Tajima's D than expected in windows of 4-

fold but not 0-fold sites (median 4fold Tajima's D = -0.533319 [95% CIs: -0.482590, -0.429297]; median 0fold Tajima's D =-0.5902975, [95% CIs: -0.606059, -0.571910]), (Sup Figure 2). Lastly, we find our observed GWA SNP set is enriched for polymorphic 0-fold

585   degenerate coding sites relative to the empirical null distribution (proportion of 0-fold sites = 0.0925%, [95% CI of the proportion of 0-fold sites per GWAS itt = 0.0137, 0.0728]).
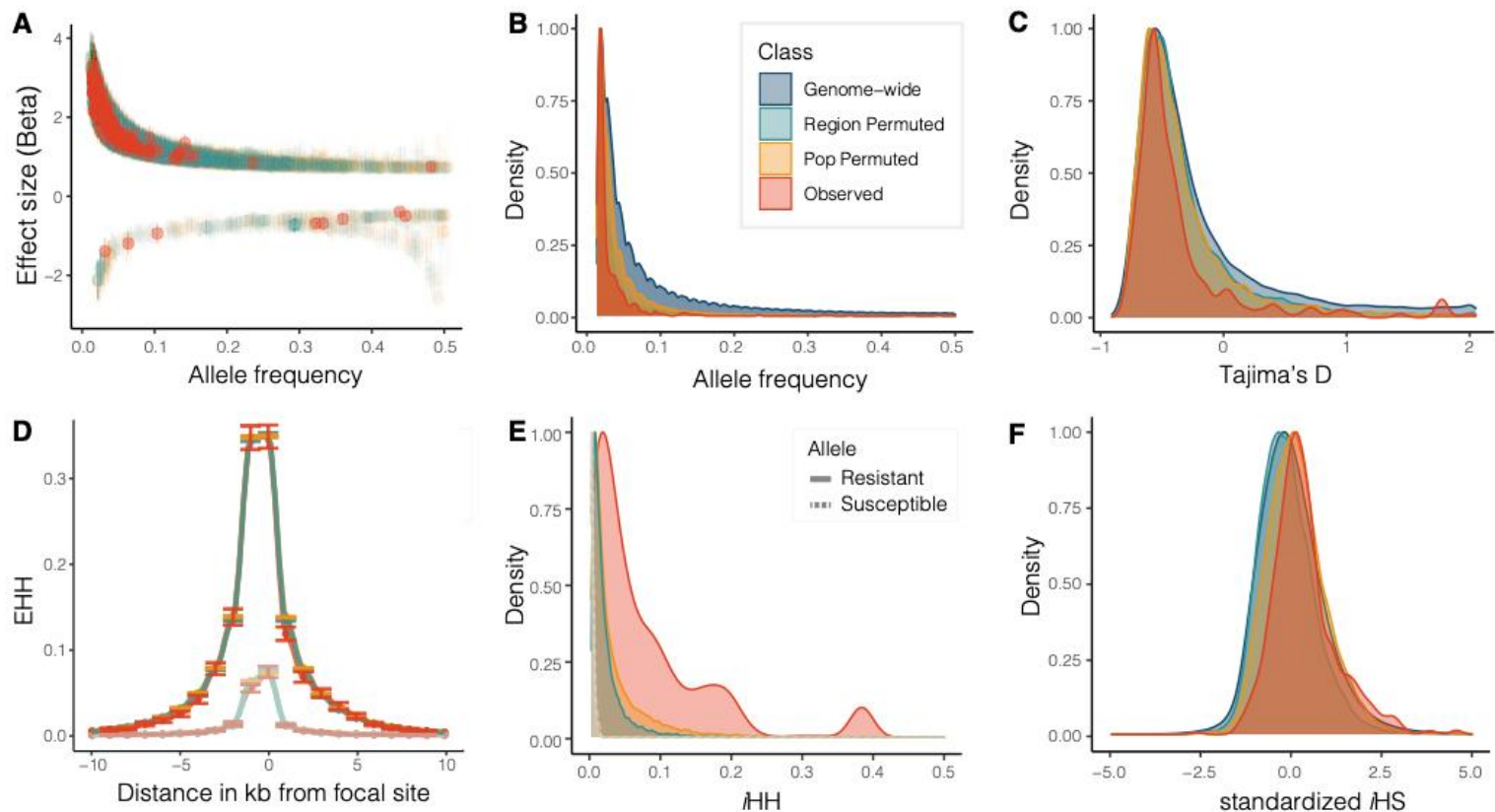

590

595

600

Figure 4. Genomic summaries of SNPs significantly associated with NTSR 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 Tajima's D, D)
605 expected haplotype homozygosity (EHH) around focal resistance-related and susceptibility-related alleles, E) the integrated haplotype homozygosity of resistant-related versus susceptibility-related haplotypes, F) standardized integrated haplotype score (iHS), treating resistance 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 B
610 applies to all plots, while legends for D and E additionally distinguish allele type.

Discussion

615

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

620 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

22

problematic agricultural weed, *Amaranthus tuberculatus.* As proof of concept, GWA correctly identified *EPSPS,* the target gene of glyphosate herbicides, and many significant SNPs neighbouring this region as significantly associated with resistance. This suggests considerable

625   LD around EPSPS, consistent with previous findings of a strong selective sweep associated with a ~6 Mb region that is amplified (Kreiner et al., 2019). Consistent with polygenic NTSR being widespread across populations and the genome, however, we find that after controlling for variation in resistance explained by TSR loci, SNPs in more than 250 distinct genes across all 16 chromosomes are also associated with glyphosate resistance.

630

Compared to both the genome-wide background and null-expectation, we find that NTSR-related alleles are enriched for particularly low frequencies and large effects. Together, these results fit a scenario of a genetic architecture of NTSR shaped by recent selection but also strong pleiotropic costs, with the involvement of specific NTSR-related alleles being highly heterogeneous across

635   populations and genetic backgrounds. Below, we discuss the genetic architecture of glyphosate resistance, the role of selection in shaping it, and candidate NTSR genes.

Model comparison, importance of NTSR versus TSR mechanisms, and population structure

With previous knowledge of two genetic causes underlying glyphosate resistance—a non-

640   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 control for variance in resistance explained by TSR to identify genome-wide SNPs contribution to polygenic, NTSR (Table 1). Reassuringly, the covariate free model showed enrichment for significant SNPs on the scaffold harbouring

645   *EPSPS* (with 32 SNPs within the *EPSPS*-amplification significantly related to glyphosate resistance and ~20% of significant GWA SNPs located on this scaffold), while controlling for TSR mechanisms led to the *EPSPS*-containing scaffold having a similar number of significant hits as an average chromosome (Fig 2).

650   Our data encompassed a set of 155 individuals where 52% are phenotypically resistant, 79% of which harboured at least one large effect TSR mechanism. That 25% of phenotypic resistance can be explained by polygenic NTSR in this sample where large-effect TSR mutations are so

23

widespread, explaining a comparable 33% of variance in resistance, speaks to the

underappreciated importance of polygenic NTSR—which may not only confer resistance in the

655 absence of TSR, but may also act as modifiers of resistance in the presence of TSR. While this

pronounced importance of polygenic resistance we describe here may even be a conservative

estimate if highly resistant individuals that lack TSR mechanisms are more widespread in the

wild than represented in our sampling, underpowered GWAs are known to overestimate the

PVE/"chip heritability" (King & Long, 2017), a source of bias which we cannot rule out here.

660 Furthermore, given our estimate of the heritability of resistance being 0.83, it is evident that the

variance explained by TSR (0.33) and NTSR mechanisms (0.25) does not fully resolve the

genetic basis of resistance.

Among populations, the importance of NTSR and TSR mechanisms are bound to vary as the

665 involvement of these genetic mechanisms differ across populations, geographic ranges, and

species (see Gaines et al., 2020 for a thorough review). This is evident in our results (Fig 1),

where the mechanisms and presence of TSR vary considerably even across nearby populations.

When focal phenotypes, or its genetic architecture, differ among populations in a manner related

to genome-wide and fine-scale relatedness among populations, alleles that come out as

670 significantly related to the trait of interest may be confounded with shared evolutionary history.

A relationship between population genetic structure, genomic background, and the genetic

mechanism(s) of a trait are a key limitation inherent to most GWAs that sample across a diverse

set of populations, in that this signal may go uncontrolled by a relatedness matrix (discussed in

Atwell et al., 2010). With this limitation in mind, we examined the extent that the genetic

675 architecture of resistance differed among populations in a manner that might be confounded with

population structure.

To do so, we took three increasingly rigorous approaches. We (i) permuted the GWA 1000x,

randomizing phenotypes and covariates with respect to genotypes globally among all samples,

680 addressing the problem of circular downstream population genomic analyses by having an

empirical null that should be equally affected by such circularities. However, to do so in a way

that can shed light on the extent to which the genomic architecture of non-target site resistance

may be confounded with population structure, we (ii) performed these permutations within

24

geographic regions that were previously found to be the main axes of differences in population

685   structure in this dataset (Kreiner et al., 2019), and (iii) performed an even more rigorous

permutation within populations. While the global permutations find on average many fewer

significant SNPs than our observed set (55/573), our regional permutations and population

permutations increasingly find a large proportion of SNPs significantly related to NTSR even

under these randomized set of phenotypes (165 and 229/573, respectively). Thus, it is evident

690   that residual population structure (unaccounted for by our relatedness matrix) and the genetic

architecture of resistance are at least partially confounded, as we obtained a large number of

significant SNPs even when phenotypes were randomly assigned to genotypes.

To gain some intuition behind these within population permutation results, consider a situation in

695   which resistance phenotype, TSR genotypes, and SNP identity at a hypothetical NTSR site in the

genome are highly correlated, or even where there is only 1 or limited combinations of genotypes

or genomic backgrounds in a population. In these instances, randomly assigning phenotypes to

different genomic backgrounds may not change the p-values or test statistic of a GWA, as each

phenotype is being randomly assigned to the same multilocus genotype, albeit from a different

700   individual. Consequently, under many permutations, we will observe p-values or test-statistics

equal to the original observation, despite random permutation, and these statistical observations

may reflect loci that have experienced selection from herbicides in relatively homogeneous

populations. Alternatively, polygenic NTSR-based resistance can become associated with

genome-wide SNPs that tag local population structure, even if they are not causally associated

705   with resistance.

While our candidate genes will need to be carefully probed in other sets of populations and with

other experimental approaches, the set of NTSR-related alleles we describe here show

particularly extreme signals of selection and herbicide resistance-related biological functions,

710   even against the most stringent population-permuted empirical null model that accounts for

uncontrolled population structure. Our results speak to widespread nature of polygenic resistance

in *Amaranthus tuberculatus* agricultural populations in North America—not just supplementing

low levels of protection against a background of large-effect TSR mechanisms, but providing

near equivalent provision of protective effects against glyphosate. Further sampling would also

715   provide an interesting comparison of the extent to which the genetic architecture of NTSR is population-specific, and how well components of our analysis are reproduced in a set of samples with markedly different population structure and possibly different architectures of resistance (as has been discussed in Josephs, Berg, Ross-Ibarra, & Coop, 2019).

720   *Effect size & Allele frequencies*
       The effect sizes of SNPs associated with NTSR tend to be negatively correlated with allele frequency (**Fig 4A**). The lack of rare alleles with small effects is likely to be driven by a 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,
725   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 2.71, while for common alleles (=>0.05) it was only 1.10 (only 57/573 significant SNPs). These beta values can be interpreted as the effect on resistance, per changing the count of the
730   allele by 1. 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
735   general lack of intermediate or common, large effect alleles may be a consequence of selection. For human diseases and gene expression in *Capsella grandiflora* and in maize, 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; Kremling et al., 2018; Kryukov et al., 2007;
740   Marouli et al., 2017). The enrichment for large effect alleles is likely to reflect particularly strong selection in shaping the genetic architecture for herbicide resistance (Eyre-Walker, 2010)). With herbicide resistance, selection for resistance approaches near lethality (s ~ 1), and thus the trade-off between these large effect alleles being advantageous on the short-term but associated with potentially long-term pleiotropic costs may contribute to keeping such alleles globally rare.

745

Similarly, the range in allele frequency and enrichment of rare alleles also implies that while a small subset of resistance-related alleles are shared among populations, most appear to be population specific. The mean frequency of NTSR alleles is 0.03 in our samples, implying such alleles are found in only 4-5 individuals. Rare population-specific alleles are likely to be the

750 work of compensatory mutations arising on already resistant backgrounds (as is the case for antibiotic resistance; Craig MacLean, Hall, Perron, & Buckling, 2010), that interact epistatically to strengthen resistance to herbicides or mediate their costs. The stacking of relatively small-effect alleles as an important genetic mechanism of NTSR has been hypothesized many times (Délye, 2013; Gressel, 2009; Kreiner et al., 2018) and rapid evolution of resistance by diverse

755 genetic mechanisms in just a few generations of selection have provided support for such a polygenic basis (Busi & Powles, 2009; P. Neve & Powles, 2005). Our work, finally provides some of the first evidence of the frequencies and effect sizes of NTSR alleles across populations, and that when combined, such alleles may provide higher resistance than TSR mechanisms.

760 *Candidate genes*

We found 573 NTSR-associated SNPs, corresponding to 274 unique *Amaranthus tuberculatus* genes. Of these genes, we could annotate 150 using orthology-based approaches. We found our gene set had diverse roles in biological processes and molecular functions, with 13 GO terms significant after FDR correction. Enriched biological processes were highly relevant not to just

765 the three-phase schema of non-target site resistance (Gaines et al., 2020)—detoxification, conjugation, and transport—but to downstream pathways known to be modulated by the shikimate pathway targeted by glyphosate. Likely our most convincing evidence that the genetic architecture we describe here is related to glyphosate usage comes from the fact that our most significantly enriched GO category is response to chemicals. The genes involved in response to

770 chemicals are diverse in their function playing roles in other enriched classes such as biosynthetic processes, response to abiotic stimulus, response to oxygen-containing compounds, and several types of carboxylic acid metabolism. Importantly, these enrichment results are much more extreme than is seen in any of within-population GWA permutations.

775 In particular, a recent analysis of the transcriptomic response to glyphosate in genetically modified soybean with one or multiple stacked TSR mechanisms (Zanatta, Benevenuto, Nodari,

& Agapito-Tenfen, 2020) provides an interesting case study, as the genetic architecture of NTSR we identify is predominantly on the background of two large-effect TSR mechanisms, akin to their genetically modified stacked-TSR backgrounds. The KEGG pathways in soybean with the

780   largest number of genes involved in response to glyphosate were metabolic pathways and biosynthesis of secondary metabolites (Zanatta et al., 2020). In comparison, of our 6 GO categories that were significant even after conservative Bonferroni correction, 4 were related to metabolism or biosynthesis. In addition to these broad metabolic pathways, our NTSR genes function in carbon-, and oxygen-related metabolic processes. Increased carbon-flux and thus

785   carbon-related metabolism may compensate for reduced levels of aromatic amino acid production in response to EPSPS inhibition by glyphosate (Maeda & Dudareva, 2012).

The synthesis of several natural products, including pigments and hormones, directly require intermediates from the shikimate pathway (Maeda & Dudareva, 2012); similar to Zanatta et al.,

790   we find that genes functioning in response to chemicals are involved in diverse regulation of auxin, ethylene, abscisic acid, jasmonic acid, and gibberellin. Since glyphosate accumulation occurs in active metabolic tissues (Cakmak, Yazici, Tutus, & Ozturk, 2009), the primary location of plant hormone production, part of the toxicity induced by glyphosate is known to be the hormonal imbalance that influences both growth and development (Gomes et al., 2014).

795   Relatedly, our genes were enriched for function in multicellular organismal processes and development, with roles in dormancy, germination, photomorphogenesis and flowering time, several of which also being enriched in response to herbicides in the single-transgene soybean variety (Zanatta et al., 2020). Together, our results suggest a key role of pleiotropic constraints, in the form of trade-offs with growth and reproduction, mediating the frequency of alleles

800   contributing to the polygenic bases of glyphosate resistance.

While pleiotropy may be one explanation for why we see genes with diverse functions in our GWA for glyphosate resistance, and this would not be unexpected given our understanding of the organismal-level impacts of herbicides (Comont et al., 2019; Dyer, 2018; Kuester, Fall,

805   Chang, & Baucom, 2017), a complex history of compounded selection from various herbicides and related shifts in life history optima could also contribute to 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). GWA studies are inherently correlative, with traits (response

810   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, 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

815   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 GWA hits. Second, any agricultural practice (pesticide use, manuring, tilling, or irrigation), soil characteristic, or climate variable that is correlated with glyphosate usage and

820   thus resistance ratings 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.

825   While we would argue that the genetic architecture of NTSR we characterize here as having convincing and direct function in response to and compensation of glyphosate toxicity, the interpretation of these resistance-related alleles as strictly driven by selection only from glyphosate herbicides is likely incorrect. These putative NTSR alleles are likely to characterize a complex history of agricultural selection in resistant weeds, and if these regimes differ among

830   populations, we expect this complex genetic architecture to be confounded with population structure. 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

835   architectures.

Haplotype homozygosity, and a role for recent positive selection

Our set of resistance-related alleles was enriched for large effect sizes and low allele frequencies (Fig 4A,B). Furthermore, we found enrichment for particularly long haplotypes associated with
840   resistance given their allele frequency (Fig 4D-F), particularly negative values of Tajima's D, and a high proportion of amino-acid changing SNPs. These particularly extreme values allowed us to interpret the history of selection on the genetic architecture of NTSR, as it is above and beyond what the permuted expectation that accounts for uncontrolled population structure in both phenotype and genotype.

845

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
850   many generations (spatially and/or temporally fluctuating selection). Weed science research has long been interested in the costs of herbicide resistance mutations (Lenormand, Harmand, & Gallet, 2018), and work on the topic has led to sometimes conflicting conclusions, 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
855   primarily on TSR mechanisms, and pleiotropic costs of resistance as a trait should be much more ubiquitous where hundreds of diverse genes are likely to be 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
860   10 kb around significant hits in our observed NTSR-related alleles compared to the null expectation and alternative susceptible haplotypes (Fig 4D). However, both iHH (integrated haplotype homozygosity, representing the space under the EHH curve) **(Fig 4E)** and a further standardized version of this statistic, the standardized integrated haplotype score (iHS; Fig 4F), are much larger than expected relative to the permuted null expectation. Since iHS is
865   standardized by genome-wide empirical allele frequency distributions, this can be interpreted as unusually large homozygous haplotypes associated with resistance relative to susceptible alleles

(Fig 3E) (Voight et al., 2006). Particularly long haplotypes associated with resistance are 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,

870 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 over time due to more opportunity for recombination (B. Charlesworth, Nordborg, & Charlesworth, 1997; D. Charlesworth, 2006; Hudson & Kaplan, 1988). That we see an excess of both standardized iHSs and resistant iHH values implies that resistance alleles have

875 been subject to positive selection. Furthermore, these haplotype-based analyses were consistent with enrichment for particularly low Tajima's D in windows, and enrichment of amino-acid changing SNPs. Taken together, base-pair-, haplotypic-, and window-based summary statistics point to the genetic architecture of NTSR having been the subject of recent positive selection, constrained by pleiotropic trade-offs.

880

Early population genetic theoretical studies have examined the effects of rotating crop and herbicide use (a rotation scenario of one year on/one year off) on herbicide resistance, assuming selection coefficients (s) for resistance between 0.75 and 0.99 within generations, and a cost of resistance in the absence of herbicides as $s=0.25$. Under such scenarios, a rare monogenic

885 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). Our results illustrate that the genetic architecture of herbicide resistance in our system is more complex, with a polygenic basis rather than a single dominant locus, and with a range of allele frequencies and effect sizes. This complex basis means that small changes in the frequency of many alleles can appreciably alter

890 phenotypic values, with weaker selection on each allele individually required. Compounded with a role of pleiotropic constraints, diversity in polygenic resistance-related alleles may persist in populations over much longer timescales then initially hypothesized.

Conclusion

895

In conclusion, our study identifies over 500 SNPs in more than 250 genes across all 16 chromosomes associated with non-target site resistance in *Amaranthus tuberculatus*. These

31

alleles are associated with large effect sizes but low frequencies, indicating a heterogeneous architecture of NTSR across populations. Consistent with the literature, we find NTSR genes

900 function in pathways directly downstream of the shikimate pathway targeted by glyphosate, but we also find that these alleles have pleiotropic roles in development and reproduction. We find particularly long haplotypes and low diversity associated with NTSR alleles, providing evidence for recent positive selection on this polygenic trait. Given that NTSR can explain 25% of the variance in phenotypic resistance, our findings imply that one can remain optimistic about the

905 prospects for phenotypic prediction of resistance in weed populations, and that even for individuals harbouring multiple large-effect TSR mutations, NTSR may be of primary importance. Moreover, our findings suggest that selection from herbicides may have more widespread consequences on genomic diversity than initially assumed. Further work to functionally validate the candidate genes identified here will shed light on mechanism and

910 consistency of these alleles in conferring resistance across populations.

Data Accessibility
- DNA sequences available at European Nucleotide Archive (ENA) (project no. PRJEB31711); https://www.ebi.ac.uk/ena/browser/view/PRJEB31711
915 - Reference genome and its annotation available at CoGe (reference ID = 54057); https://genomevolution.org/coge/GenomeInfo.pl?gid=54057
- Scripts for data analysis available at https://github.com/jkreinz/NTSR-GWAS-MolEcol2020.
- Sampling locations and morphological data available in Sup Table 1.
920

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.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.

Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., … Nordborg, M. (2010). Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature*, *465*(7298), 627–631.

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

Brewer, P. B., Howles, P. A., Dorian, K., Griffith, M. E., Ishida, T., Kaplan-Levy, R. N., … Smyth, D. R. (2004). PETAL LOSS, a trihelix transcription factor gene, regulates perianth architecture in the Arabidopsis flower. *Development* , *131*(16), 4035–4045.

Busi, R., & Powles, S. B. (2009). Evolution of glyphosate resistance in a Lolium rigidum

960   population by glyphosate selection at sublethal doses. *Heredity*, *103*(4), 318–325.

Cakmak, I., Yazici, A., Tutus, Y., & Ozturk, L. (2009). Glyphosate reduced seed and leaf concentrations of calcium, manganese, magnesium, and iron in non-glyphosate resistant soybean. *European Journal of Agronomy: The Journal of the European Society for Agronomy*, *31*(3), 114–119.

965   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.

970   Comont, D., Knight, C., Crook, L., Hull, R., Beffa, R., & Neve, P. (2019). Alterations in Life-History Associated With Non-target-site Herbicide Resistance in Alopecurus myosuroides. *Frontiers in Plant Science*, *10*, 837.

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

975   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.

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

980   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

985   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.

990   Duke, S. O., & Powles, S. B. (2008). Glyphosate: a once-in-a-century herbicide. *Pest*

34

Management Science: Formerly Pesticide Science, *64*(4), 319–325.

Dyer, W. E. (2018). Stress-induced evolution of herbicide resistance and related pleiotropic effects. *Pest Management Science*, *74*(8), 1759–1768.

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.

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

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.

Gressel, J. (2009). Evolving understanding of the evolution of herbicide resistance. *Pest Management Science*, *65*(11), 1164–1173.

35

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)*.

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

Ito, H., & Gray, W. M. (2006). A gain-of-function mutation in the Arabidopsis pleiotropic drug resistance transporter PDR9 confers resistance to auxinic herbicides. *Plant Physiology*, *142*(1), 63–74.

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., Berg, J. J., Ross-Ibarra, J., & Coop, G. (2019). Detecting Adaptive Differentiation in Structured Populations with Genomic Data and Common Gardens. *Genetics*, *211*(3), 989–1004.

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.

King, E. G., & Long, A. D. (2017). The Beavis Effect in Next-Generation Mapping Panels in Drosophila melanogaster. *G3* , *7*(6), 1643–1652.

Kim, S., Choi, H.-I., Ryu, H.-J., Park, J. H., Kim, M. D., & Kim, S. Y. (2004). ARIA, an Arabidopsis arm repeat protein interacting with a transcriptional regulator of abscisic acid-responsive gene expression, is a novel abscisic acid signaling component. Plant Physiology,

36

136(3), 3639–3648.

Kreiner, Giacomini, Bemm, Waithaka, Regalado, Lanz, Hildebrandt, Sikkema, Tranel, Weigel,
1055   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,
1060   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.
1065   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–
1070   635.

Kremling, K. A. G., Chen, S.-Y., Su, M.-H., Lepak, N. K., Romay, M. C., Swarts, K. L., …
Buckler, E. S. (2018). Dysregulation of expression correlates with rare-allele burden and
fitness loss in maize. *Nature*, *555*(7697), 520–523.

Kryukov, G. V., Pennacchio, L. A., & Sunyaev, S. R. (2007). Most rare missense alleles are
1075   deleterious in humans: implications for complex disease and association studies. *American
Journal of Human Genetics*, *80*(4), 727–739.

Kuester, A., Fall, E., Chang, S.-M., & Baucom, R. S. (2017). Shifts in outcrossing rates and
changes to floral traits are associated with the evolution of herbicide resistance in the
common morning glory. *Ecology Letters*, *20*(1), 41–49.

1080   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.

Kurepa, J., Smalle, J., Va, M., Montagu, N., & Inzé, D. (1998). Oxidative stress tolerance and

1085   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.

1090   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.

1095   Lenormand, T., Harmand, N., & Gallet, R. (2018). Cost of resistance: an unreasonably expensive concept. *Rethinking Ecology*, Vol. 3, pp. 51–70. doi: 10.3897/rethinkingecology.3.31992

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.

1100   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.

Lu, W., Tang, X., Huo, Y., Xu, R., Qi, S., Huang, J., … Wu, C.-A. (2012). Identification and characterization of fructose 1,6-bisphosphate aldolase genes in Arabidopsis reveal a gene

1105   family with diverse responses to abiotic stresses. *Gene*, Vol. 503, pp. 65–74. doi: 10.1016/j.gene.2012.04.042

Maeda, H., & Dudareva, N. (2012). The shikimate pathway and aromatic amino Acid biosynthesis in plants. *Annual Review of Plant Biology*, *63*, 73–105.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., …

1110   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.

1115   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
1120   https://pdfs.semanticscholar.org/37c0/71c2bcd115bcce4bb2e4eb89a7dbb861c136.pdf

        Mhamdi, A., Mauve, C., Gouia, H., Saindrenan, P., Hodges, M., & Noctor, G. (2010). Cytosolic
        NADP-dependent isocitrate dehydrogenase contributes to redox homeostasis and the
        regulation of pathogen responses in Arabidopsis leaves. *Plant, Cell & Environment*, *33*(7),
        1112–1123.

1125   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.

1130   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.

        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
1135   Altered Target-Site and Nontarget-Site Mechanisms. *Weed Science*, *61*(3), 374–383.

        Neve, P., & Powles, S. (2005). High survival frequencies at low herbicide use rates in
        populations of Lolium rigidum result in rapid evolution of herbicide resistance. *Heredity*,
        95(6), 485–492.

        Neve, P., Vila-Aiub, M., & Roux, F. (2009). Evolutionary-thinking in agricultural weed
1140   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.

        Pandey, S. P., & Somssich, I. E. (2009). The role of WRKY transcription factors in plant
        immunity. *Plant Physiology*, *150*(4), 1648–1655.

1145   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.

1150 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.

Phukan, U. J., Jeena, G. S., & Shukla, R. K. (2016). WRKY Transcription Factors: Molecular Regulation and Stress Responses in Plants. *Frontiers in Plant Science*, *7*, 760.

1155 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),
1160 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
1165 pyrosequencing and its application for studies of herbicide target-site genes. *Pest Management Science*, *66*(10), 1042–1052.

Riondet, C., Desouris, J. P., Montoya, J. G., Chartier, Y., Meyer, Y., & Reichheld, J.-P. (2012). A dicotyledon-specific glutaredoxin GRXC1 family with dimer-dependent redox regulation is functionally redundant with GRXC2. *Plant, Cell & Environment*, Vol. 35, pp. 360–373.
1170 doi: 10.1111/j.1365-3040.2011.02355.x

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.

Sasaki, E., Köcher, T., Filiault, D. L., & Nordborg, M. (n.d.). *Revisiting a GWAS peak in*
1175 *Arabidopsis thaliana reveals possible confounding by genetic heterogeneity*. doi: 10.1101/2021.02.03.429533

Seo, H. S., Song, J. T., Cheong, J. J., Lee, Y. H., Lee, Y. W., Hwang, I., … Choi, Y. D. (2001). Jasmonic acid carboxyl methyltransferase: a key enzyme for jasmonate-regulated plant responses. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(8), 4788–4793.

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

Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research*, *23*(1), 23–35.

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.

Tepperman, J. M., Hudson, M. E., Khanna, R., Zhu, T., Chang, S. H., Wang, X., & Quail, P. H. (2004). Expression profiling ofphyBmutant demonstrates substantial contribution of other phytochromes to red-light-regulated gene expression during seedling de-etiolation. *The Plant Journal*, Vol. 38, pp. 725–739. doi: 10.1111/j.1365-313x.2004.02084.x

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.

Vaishak, K. P., Yadukrishnan, P., Bakshi, S., Kushwaha, A. K., Ramachandran, H., Job, N., …

    Datta, S. (2019). The B-box bridge between light and hormones in plants. *Journal of*

1210   *Photochemistry and Photobiology. B, Biology*, *191*, 164–174.

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., …

1215   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.

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.

1220 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.

Wang, Q., Zeng, J., Deng, K., Tu, X., Zhao, X., Tang, D., & Liu, X. (2011). DBB1a, involved in

    gibberellin homeostasis, functions as a negative regulator of blue light-mediated hypocotyl

    elongation in Arabidopsis. *Planta*, *233*(1), 13–23.

1225 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.

Xi, J., Xu, P., & Xiang, C.-B. (2012). Loss of AtPDR11, a plasma membrane-localized ABC

    transporter, confers paraquat tolerance in Arabidopsis thaliana: A paraquat transporter in

1230   Arabidopsis. *The Plant Journal: For Cell and Molecular Biology*, *69*(5), 782–791.

Xu, S. (2003). Theoretical basis of the Beavis effect. *Genetics*, *165*(4), 2259–2268.

Yoshida, N., Yanai, Y., Chen, L., Kato, Y., Hiratsuka, J., Miwa, T., … Takahashi, S. (2001).

    EMBRYONIC FLOWER2, a novel polycomb group protein homolog, mediates shoot

    development and flowering in Arabidopsis. *The Plant Cell*, *13*(11), 2471–2481.

1235 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

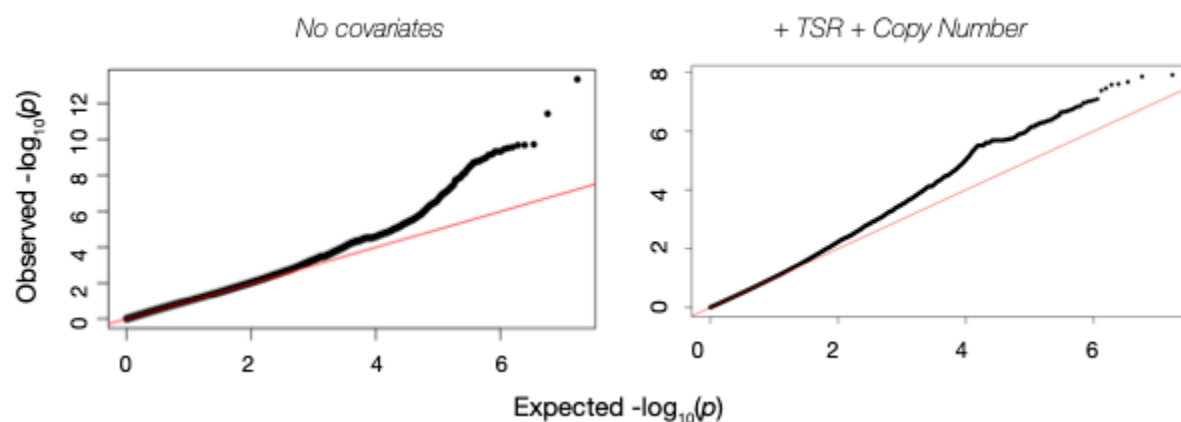   mechanisms endow resistance to glyphosate, ACCase and ALS-inhibiting herbicides in

multiple herbicide-resistant Lolium rigidum. *Planta*, *230*(4), 713–723.

1240   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.
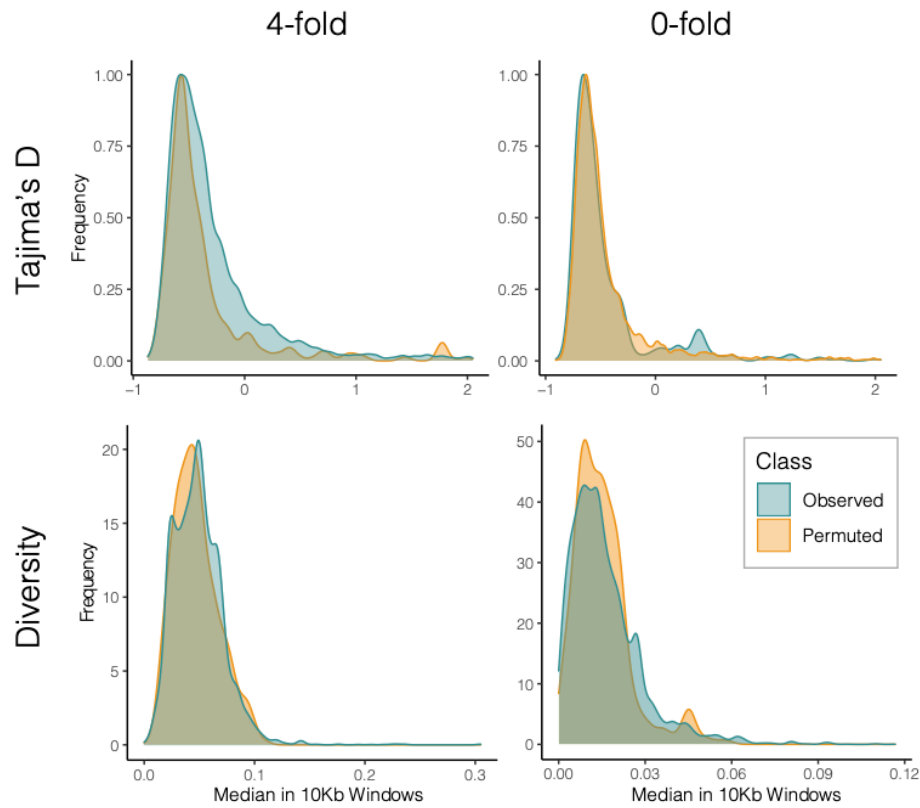
       Zanatta, C. B., Benevenuto, R. F., Nodari, R. O., & Agapito-Tenfen, S. Z. (2020). Stacked
       genetically modified soybean harboring herbicide resistance and insecticide rCry1Ac shows
       strong defense and redox homeostasis disturbance after glyphosate-based herbicide

1245   application. *Environmental Sciences Europe*, *32*(1), 104.

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

Supplementary Figures

1250

Sup figure 1. QQplots of observed versus expected p-values for GWA models of raw resistance
levels (no covariates) and a model of non-target site resistance (using the EPSPS Pro106Ser
substitution and EPSPS copy number as covariates).

1255

Sup Figure 2. Distributions of 4 fold and 0 fold Tajima's D and diversity in 10kb windows surrounding focal NTSR SNPs (across all scaffolds, or excluding the EPSPS-containing scaffold 5) relative to the permuted distributions.

1260