Title: Inter-chromosomal linkage disequilibrium and linked fitness cost loci influence the evolution of nontarget site herbicide resistance in an agricultural weed

Sonal Gupta[1*], Alex Harkess[2,3*], Anah Soble[1], Megan Van Etten[4], James Leebens-Mack[5], and Regina S Baucom[1*]

[1]Ecology and Evolutionary Biology Department, 4034 Biological Sciences Building, University of Michigan, Ann Arbor, MI 48109
[2]Department of Crop, Soil, and Environmental Sciences, Auburn University, Auburn, AL 36849
[3]HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806
[4]Biology Department, Pennsylvania State University, Dunmore, PA 18512
[5]Department of Plant Biology, University of Georgia, Athens, GA, 30602
[*]Corresponding authors

Abstract
The adaptation of weedy plants to herbicide is both a significant problem in agriculture and a model for the study of rapid adaptation under regimes of strong selection. Despite recent advances in our understanding of simple genetic changes that lead to resistance, a significant gap remains in our knowledge of resistance controlled by many loci and the evolutionary factors that influence the maintenance of resistance over time. Here, we perform a multi-level analysis involving whole genome sequencing and assembly, resequencing and gene expression analysis to both uncover putative loci involved in nontarget herbicide resistance and to examine evolutionary forces underlying the maintenance of resistance in natural populations. We found loci involved in herbicide detoxification, stress sensing, and alterations in the shikimate acid pathway to be under selection, and confirmed that detoxification is responsible for glyphosate resistance using a functional assay. Furthermore, we found interchromosomal linkage disequilibrium (ILD), most likely associated with epistatic selection, to influence NTSR loci found on separate chromosomes thus potentially mediating resistance through generations. Additionally, by combining the selection screen, differential expression and LD analysis, we identified fitness cost loci that are strongly linked to resistance alleles, indicating the role of genetic hitchhiking in maintaining the cost. Overall, our work strongly suggests that NTSR glyphosate resistance in *I. purpurea* is conferred by multiple genes which are maintained through generations *via* ILD, and that the fitness cost associated with resistance in this species is a by-product of genetic-hitchhiking.

Keywords: NTSR resistance, detoxification, cost, interchromosomal linkage disequilibrium, genetic hitchhiking

Introduction

Pesticide and herbicide use has reshaped ecological networks and induced strong selective pressures in the anthropogenic era. How species may adapt to strong selection is a fundamental question in evolution with great importance to the control of pesticide resistant organisms. A striking feature of pesticide resistance evolution is that there are a number of different genetic solutions that can lead to resistance (1, 2). In herbicide resistant plants, for example, resistance can be due to single gene mutations, often found in the herbicide's target protein (target site resistance, TSR), or due to changes in multiple genes, often underlying nontarget herbicide resistance (NTSR) mechanisms (3, 4). A growing body of work has produced a better understanding of resistance controlled by single genes across a variety of species (3, 5, 6). However, we currently lack a deep understanding of both the genetic basis and evolutionary potential of nontarget site resistance mechanisms genome wide (7–9).

This is due in part to the broad nature of nontarget site herbicide resistance mechanisms more generally. NTSR can be caused by reduced herbicide uptake or penetration, altered translocation or sequestration, and/or herbicide detoxification (10–12) -- mechanisms that likely rely on a complex genetic basis (8, 13–15). While some investigations have pinpointed a single gene conferring NTSR (16, 17), gene expression surveys or whole genome re-sequencing assays in a small handful of resistant weeds are beginning to shed light on the complexity of nontarget resistance mechanisms (18–20). For example, in both *Amaranthus tuberculatus* and *Ipomoea purpurea*, a number of different loci found across the genome -- whether structural, regulatory, or both -- exhibit signs of selection and are thus putatively involved in resistance (18–20). Because we lack a deep understanding of the genetic basis of NTSR in most weeds, however, we lack a firm grasp on the underlying forces that influence the maintenance of resistance in natural populations, such as the prevalence of alleles that may contribute to fitness costs of resistance, or the presence of interchromosomal linkage disequilibrium (ILD). The presence of ILD between unlinked regions of the genome would implicate the potential for epistatic interactions between alleles underlying either resistance or its cost.

Ipomoea purpurea is a common agricultural weed in the southeast and Midwest United States. Populations of this species, which have consistently been exposed to glyphosate based herbicides since the late 1990's (21, 22), exhibit varying levels of herbicide resistance, with some populations exhibiting low and others high survival post-herbicide application (21). There is a fitness cost associated with this resistance: resistant populations show lower germination and deteriorated seed quality compared to susceptible populations(23). Further, populations from the south and Midwest show evidence of genetic admixture, with both microsatellite and SNP data showing low genetic differentiation (F_{ST} = 0.11-0.14, (24) and recent genetic connectivity (24)). RADseq and

exome sequencing has identified regions of the genome under selection and thus associated with herbicide resistance. These regions are enriched for cytochrome P450s, glycosyltransferases, and ABC transporter genes, indicating a likely role of herbicide detoxification in conferring resistance (20). Despite evidence that detoxification underlies resistance in this species, and suggestions that loci found on different chromosomes contribute to resistance, previous work relied on low-coverage RADseq sequencing without the benefit of a contiguously assembled genome. Thus, loci that may contribute to NTSR or its cost were likely missed (25), meaning that we lack a thorough understanding of NTSR, the genomic context of NTSR alleles, and the potential for relationships among NTSR alleles in this species -- all crucial to understanding the evolution of resistance more broadly.

Here, we implemented a genome-wide selection screen using whole-genome resequencing of natural populations along with a gene expression survey to characterize the genetic architecture of glyphosate resistance and its cost in *Ipomoea purpurea*. We complemented our survey with a functional assay to test the potential that resistant *I. purpurea* individuals detoxify the herbicide. Given previous evidence that multiple loci likely contribute to herbicide resistance in this species, and evidence of fitness cost of resistance, we made two main predictions regarding genome-wide patterns of selection associated with resistance in *I. purpurea*. First, we expected that regions of the genome showing high differentiation and marks of selection when comparing herbicide resistant and susceptible individuals would contain loci with strong functional links to either herbicide resistance or its cost. Second, we anticipated that linkage disequilibrium, the non-random association of alleles at different loci, should be evident among regions of the genome housing resistance loci. Although inter-chromosomal linkage disequilibrium has been identified in other systems assessing ecologically relevant traits such as mate choice and coloration (26–28), it is unknown if loci underlying herbicide resistance that are found across chromosomes exhibit long-distance or inter-chromosomal linkage disequilibrium, as would be expected if adaptation to herbicide is facilitated by multilocus genotypes favored by selection (*i.e.*, coadapted gene complexes) (29–31).

Results

A chromosome-scale genome assembly for common morning glory

We assembled a reference *I. purpurea* genome to test these hypotheses, generating the first genome sequence for this common and noxious weed. We generated a total of 48 gigabases of PacBio Sequel whole genome shotgun data (Supplemental Figure S1a). Based on a flow cytometry genome size (Benaroya Institute, Seattle, WA), this amounts to roughly 59X genome coverage for an estimated haploid genome size of 814 Mb. We used 34.79

gigabases of trimmed and self-corrected reads for assembly, scaffolding and polishing, which produced a 602 Mb assembly in 402 scaffolds (434 contigs), with a scaffold N50 of 5.77 Mb.

We performed pseudomolecule scaffolding with Phase Genomics Hi-C map, which collapsed the assembly into the expected 15 haploid chromosomes (Supplemental Figure S1b). We renamed and oriented chromosomes according to a high degree of synteny with the related *Ipomoea nil* genome (32) (Supplemental Figure S1c). No misjoins were identified and broken based on the Hi-C linkage data. BUSCO scores on the unannotated assembly show 97.5% completeness against the Viridiplantae odb10 gene set (Supplemental Figure S1d).

Approximately 63% of the assembly was masked as repetitive DNA, with a significant proportion of recently expanded Long Terminal Repeat (LTR) retrotransposons (Supplemental Figure S1e). Given the high degree of synteny with *I. nil* genome, the discrepancy between the flow cytometry genome size (814 Mb) and the assembled size (602 Mb) is likely due to young retrotransposon proliferation. We annotated 53,973 genes by combining ab initio gene predictions and RNA sequencing data from leaf tissue. The assembly shows a high degree of synteny with several genomes in the Convolvulaceae family, including *I. nil, I. trifida, and I. triloba (*Supplemental Figure S2*).*

Detecting loci under selection

Whole-genome analysis of 69 individuals identified twenty-one regions across the genome exhibiting signals of strong genetic differentiation and signs of selection when comparing herbicide resistant and susceptible populations (Supplementary TableS4). These regions exhibited a G_{ST} and Md-rank-P in the top 5 percentile ($G_{ST} > 0.24$ and Md-rank-P > 5.10) with the mid-rank P being a composite test of selection that incorporates nucleotide diversity, Tajima's D, Fay and Wu's H, and H12. This strategy identified 6.6 Mb of the genome showing signs of selection associated with herbicide resistance. The regions under selection were located across nine chromosomes, varied in size between 24kb-1738kb (Figure 1), and housed 551 genes, 323 of which could be functionally annotated (Supplementary TableS5).

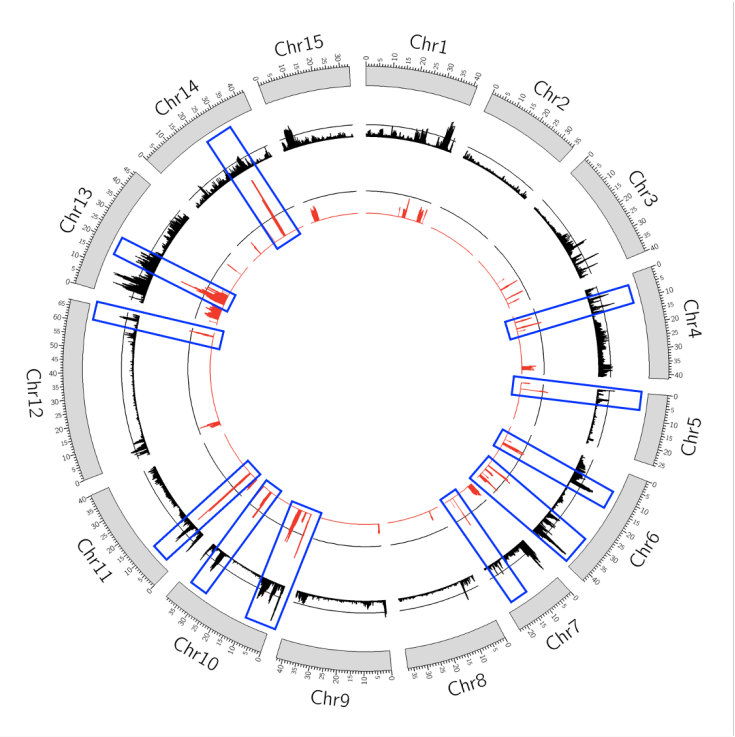


Figure 1 Circos plot depicting the regions of the genome that show signs of selection associated with herbicide resistance. The genome assembly resulted in 15 scaffolds which are represented here by grey bars. Values of G_{ST} describing the differentiation between the resistant and the susceptible populations are depicted by black bars, and the Md-rank-*P* values identifying signatures of selection are presented in red bars. Regions of the genome that exhibited both high differentiation (G_{ST} > 0.24) and a significant Md-rank-*P* value (Md-rank-P > 5.10) are identified by blue boxes. Black lines above both the G_{ST} and Md-rank-*P* represent 95% most extreme genome-wide values for each metric. Blue boxes on Chr5, Chr6, Chr10, Chr13, and Chr14 represent more than one region under selection.

The strongest signal of selection we uncovered was found within a 335kb region of chromosome 10 (average Md-rank-*P* = 7.16; average G_{ST} = 0.61, Figure 2). Within this region we identified 9 copies of cytochrome P450 genes (CYP) and 8 copies of glycosyltransferases, both of which are gene families previously implicated in herbicide detoxification. Eight of these cytochrome P450s belong to the 76A family (CYP76A1 and CYP76A2) and were present in tandem within 53kb. Four copies of the cytochrome P450s exhibited multiple non-synonymous mutations that were almost fixed in the resistant individuals (allele frequency = 0.95). Two of the eight cytochrome P450s (CYP76A2) in this block exhibited either a premature stop codon and/or a splice site donor variant (G->C) in the first intron (allele1 susceptible frequency = 0.68, resistant frequency = 0.05) in the majority of the susceptible individuals. Seven glycosyltransferases were found in tandem; one glycosyltransferase copy showed the loss of a stop codon (susceptible frequency = 0.64, resistance frequency = 0.05), whereas the other glycosyltransferases exhibited

multiple non-synonymous mutations close to fixation in the resistant individuals (resistant frequency = 0.05, susceptible frequency = 0.60; Supplementary Figure S3). Additionally, the cytochrome P450s and the block of glycosyltransferases in this region showed evidence of a hard sweep (CYP H12 = 0.88, glycosyltransferases H12 = 0.97).

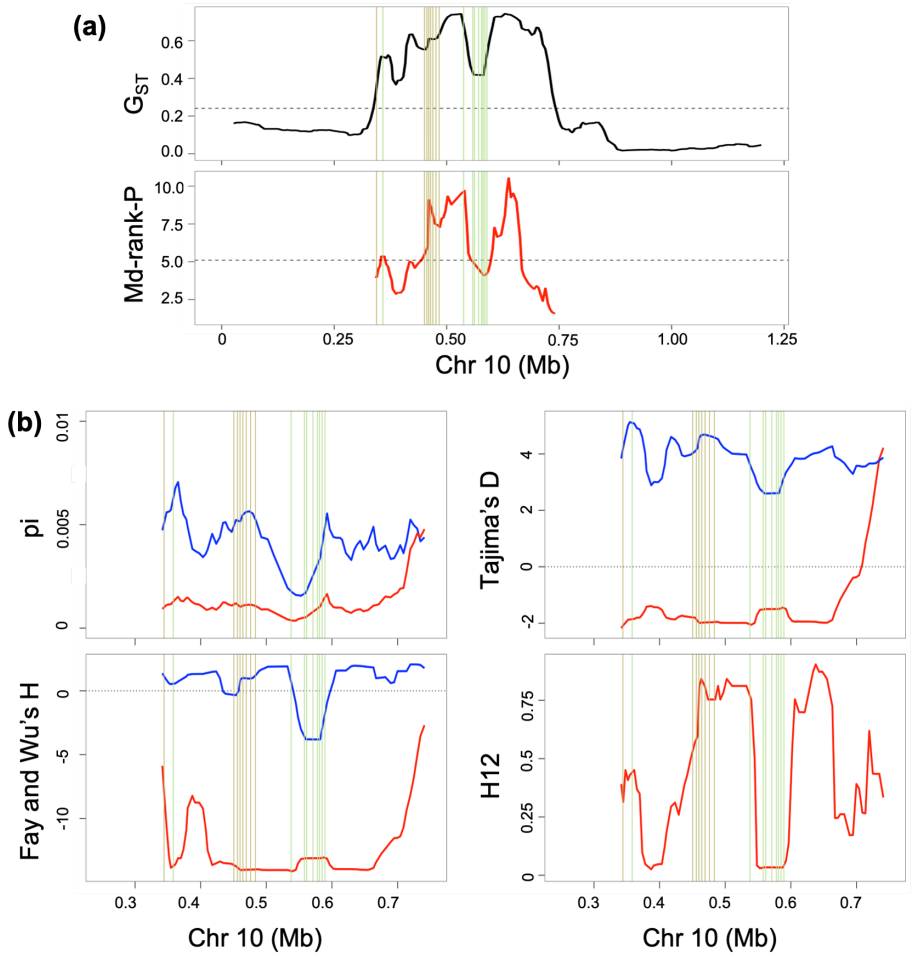


Figure 2 Region of Chromosome 10 showing signs of selection. Shown is the (a) G_{ST} (upper) and Md-rank-*P* (lower) for the resistant individuals which was estimated using statistics shown in (b) clockwise starting from upper left, pi, Tajima's D, H12 and Fay and Wu's H. Red and blue lines indicate respective values for the resistant and the susceptible populations, respectively. Khaki vertical lines represent copies of glycosyltransferases, green vertical lines are the cytochrome P450. The black dashed line in (a) represents 95 percentile values.

On chromosome 11, we found a 6.4-7.1Mb region to show signs of selection (average Md-rank-*P* = 7.709; average G_{ST} = 0.386) with six copies of a phosphate transporter gene (*PHO1*), an ABC transporter gene (*ABCB19*), and a sugar transporter gene (*ERD6*), all containing almost fixed non-synonymous mutations in resistant individuals (resistant

frequency = 0.99). This region also contained two copies each of a glycosyltransferase and a cytochrome P450 gene (CYP736A12 family, Supplementary TableS5).

Another region of note showing strong signals of selection was found on chromosome 6 (average Md-rank-*P* = 5.256; average G_{ST} = 0.762, Figure 1), with evidence of strong differentiation continuing further upstream and downstream (40.24Mb - 40.81Mb; mean G_{ST} = 0.716). Within the extended downstream region, we found ethylene responsive transcription factor (*ERF4*) and multiple copies of serine/threonine kinases, genes that are involved in the signal transduction in response to various biotic and abiotic stresses (33–37). Within this region we also identified loci that are likely related to the cost of resistance in this species, expanded upon further in 'Signs of selection on potential cost loci' below.

Across the other regions exhibiting signs of selection, we found multiple environmental stress response genes (Supplemental TableS4): *LOG3* (chr 4), associated with drought stress response (38), *MDAR4* (chr 5), involved in hydrogen peroxide detoxification (39), the GT-3B transcription factor (chr 6) which is responsible for inducing response to salt (40), and several homologs of FRS related sequence (FRS) and E3 ubiquitin-protein ligases, genes involved in response to oxidative stress (34, 37). Of special note, we uncovered two genes involved in the shikimate acid pathway -- on chromosome 10, the bifunctional 3-dehydroquinate dehydratase/shikimate dehydrogenase (DHD/SHD) gene, which is responsible for converting dehyroquinate to shikimate (41), and on chromosome 14, the arogenate dehydrogenase 2 (*TYRAAT2*) gene, which is involved in the ultimate steps of tyrosine biosynthesis (42). These latter two genes are notable in that they comprise parts of the shikimic acid pathway, which is the biochemical pathway inhibited by glyphosate (43).

Overall, our selection screen using a WGS resequencing approach identified highly differentiated regions under selection, with these regions containing genes involved in herbicide detoxification (cytochromeP450s, glycosyltransferases, ABC and phosphate transporters), the shikimate acid pathway (DHD/SHD and *TYRAAT2*), environmental sensing (serine/threonine kinases, FBL2, RAV2, and PLT1), and stress response genes (ERFs, E3 ubiquitin-protein ligase, FRSs, AHL, MSRB5, LOG3, GT-3B transcription factor). Thus, our study expands on our previous work which found detoxification genes to be under selection (20) by providing strong evidence that glyphosate resistance in *I. purpurea* is controlled by a polygenic NTSR mechanism likely involving herbicide detoxification, response to environmental stimuli and stress, and components of the shikimate acid pathway, which is the biochemical pathway inhibited by glyphosate.

Gene expression differences implicate herbicide detoxification

We compared gene expression between herbicide treated resistant and susceptible plants and found support for the idea that herbicide detoxification, plant signalling, and stress

response underlies resistance. Of the 250 differentially expressed genes (111 upregulated and 139 downregulated; Supplementary TableS6), we found cytochrome P450s, glycosyltransferases, and glutathione S-transferase genes (Figure 3a) to be differentially regulated between resistant and susceptible plants. Two copies of the cytochrome P450 family CYP82D7 were significantly upregulated in the resistant individuals (logFC: 2.05 and 1.35), along with two copies of UDP-glycosyltransferases (UGT87A2 and UGT88B1) and a glutathione S-transferase (GST). We additionally found a cytochrome P450 (CYP82C4) and a glycosyltransferase (UGT89B2) downregulated in the resistant individuals.

We likewise uncovered differences in the expression of genes associated with environmental stress responses. Among notable genes were ethylene responsive transcription factors (ERF003, ERF107, TINY; (44, 45)), serine/threonine kinase BLUS1 (46), E3 Ubiquitin protein ligase PUB23 (34), NAC domain containing protein 72 (47), and WRKY transcription factors (WRKY4, WRKY31, WRKY75; (48–50)) (Figure 3a). Homologs of these genes (ERF4, PLT1, CRT1, PRP4, HT1, B120, RHC1A, RF298, NAC92, NAC25, WRKY22; Supplementary Table S5) were also under selection when comparing herbicide resistant and susceptible populations.

In the control (non-herbicide) environment, we found 623 differentially expressed genes when comparing resistant and susceptible individuals (319 upregulated and 304 downregulated; Supplementary TableS7). We identified multiple copies of cytochrome P450s, glycosyltransferases, and ABC transporters that were differentially expressed, indicating that glyphosate resistance through detoxification is constitutive, and not induced, in this species. Interestingly, the specific cytochrome P450s and glycosyltransferase genes that exhibited signs of selection from our whole-genome scan were not the same as those that exhibited differential expression, which could be due to the non-simultaneous nature of the gene transcription response to glyphosate (51), or could represent a transcriptional sampling stage caveat.

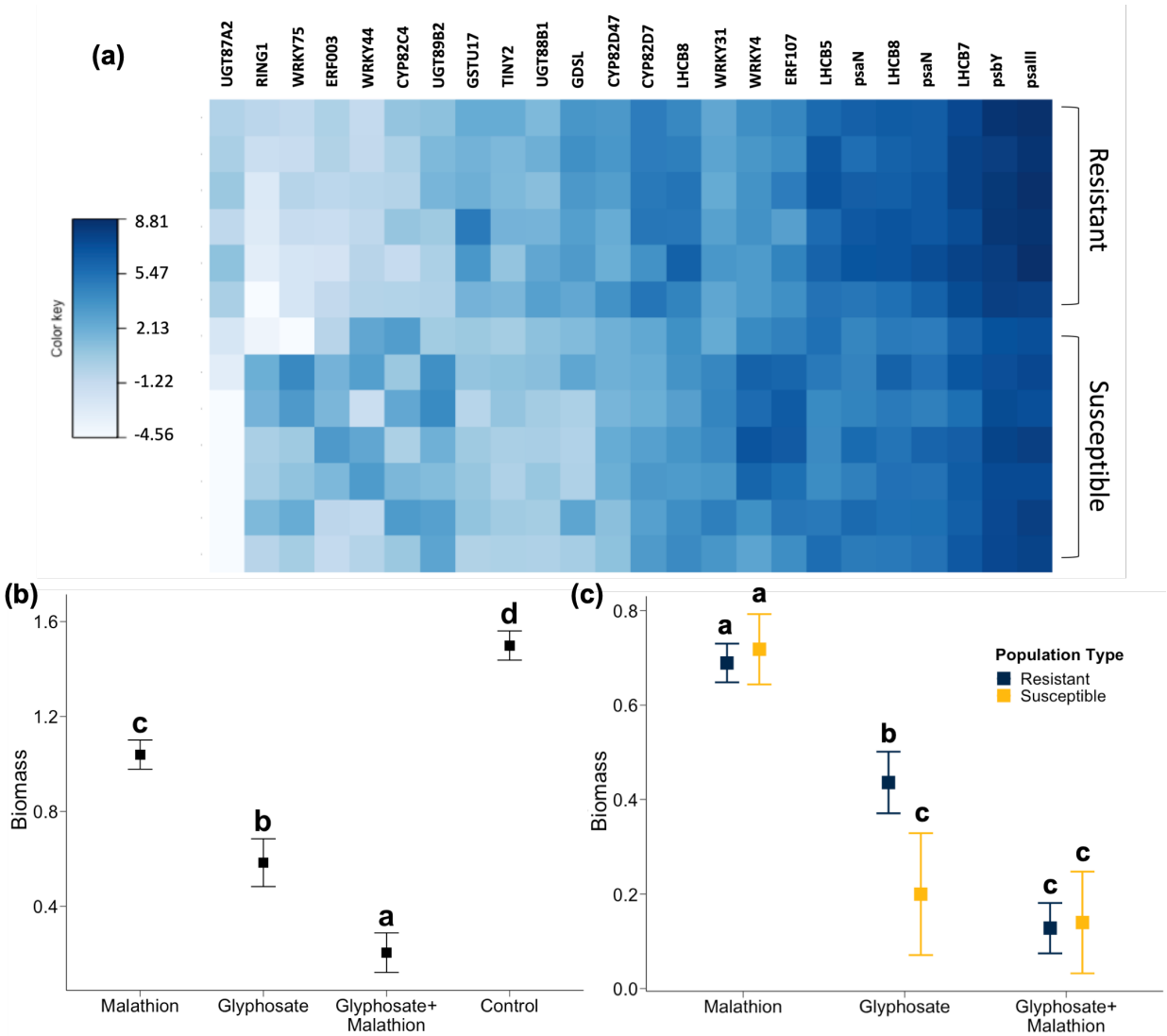


Figure 3 Gene expression variation associated with herbicide resistance, and results of a functional assay supporting the idea that resistance in *I. purpurea* is due to detoxification. (a) Loci associated with glyphosate resistance identified by differential expression analysis with P-value < 0.0005. Color key represents log2 fold-change values. (b) Least square means of above-ground biomass according to treatment (malathion, glyphosate, glyphosate plus malathion, and a control (no treatment)) and (c) summarized according to resistance type (R/S). Letters in (b) and (c) indicate significant differences between treatment environments. The addition of the cytochrome P450 inhibitor malathion reverses glyphosate resistance (glyphosate vs glyphosate+malathion, contrast estimate = 0.379, t-ratio = 2.946, p-value = 0.019), with the resistant individuals showing the same phenotype as the susceptible individuals in the presence of glyphosate and malathion but not in the presence of glyphosate only.

Functional assay supports herbicide detoxification as a mechanism of resistance

We performed an assay to determine if the functional mechanism of resistance in *I. purpurea* was herbicide detoxification (following 17, 52–55). We applied malathion, a pesticide that inhibits cytochrome P450s, to multiple resistant and susceptible *I. purpurea* individuals from the same populations used in the WGS re-sequencing and gene expression studies. The expectation that malathion would act to inhibit *I. purpurea* cytochrome P450s was met; we found a significant overall treatment effect (F-value = 59.33, df = 3, $p < 0.0001$; Fig 3b) with individuals treated with both glyphosate and malathion showing lower biomass compared individuals treated with either malathion, glyphosate, or untreated controls (Fig 3b; Supplemental TableS8).

As expected, resistant individuals showed significantly greater biomass compared to the susceptible individuals in the presence of glyphosate (F-value = 4.81, df = 1, P-value = 0.03; Figure 3b). However, the biomass of resistant individuals in the presence of both malathion and glyphosate was significantly lower than that of resistant individuals treated only with glyphosate (resistant plants, malathion+glyphosate *vs* glyphosate: t = 3.65, df = 78, p-value = 0.001), indicating that the presence of malathion reduces the resistance response. In fact, the presence of both malathion and glyphosate led to similar (and low) remaining biomass of both resistant and susceptible individuals (malathion+glyphosate treatment: resistant vs susceptible plants: t = 0.15, df = 32, p-value = 0.88). This shows that the presence of a cytochrome P450 inhibitor lowers the level of glyphosate resistance in *I. purpurea* plants, supporting the idea that modification to the detoxification pathway underlies glyphosate resistance in this species.

Role of long-distance and interchromosomal linkage disequilibrium in maintaining NTSR alleles

Our whole-genome scan identified regions under selection containing genes involved in environmental sensing, stress responses, and herbicide detoxification. This broad scan implicates a polygenic basis of resistance in *I. purpurea* and shows that multiple regions of the genome likely contribute to resistance. We thus sought to determine if there was evidence of linkage disequilibrium between these regions, which would potentially suggest either epistatic interactions among alleles or the inheritance of coadapted gene complexes (29, 56). We calculated a measure of linkage disequilibrium (r^2) between long-distance and interchromosomal SNPs that showed the most extreme level of differentiation (98th percentile, $G_{ST} > 0.35$) -- regions on Chromosome 4, 6 (two regions, from hereon referred to as 6.1 and 6.2), 10 and 11, and compared it to the whole-genome measure. We found that the five regions under selection showed islands of elevated interchromosomal linkage disequilibrium (Supplementary TableS9) in a backdrop of nearly zero genome-wide ILD (background interchromosomal r^2 mean = 0.00096; Figure 4). The region under selection on Chr10 exhibited the strongest linkage to other chromosomal regions under selection

(99% ILD Chr4-Chr10 = 0.256, Chr6.1-Chr10 = 0.262, Chr6.2-Chr10 = 0.22, Chr11-Chr10 = 0.165). Interestingly, the max r^2 values within these regions was observed for putative resistance genes identified above. For instance, multiple glycosyltransferases under selection on Chr10 showed max ILD with SNPs on Chr11 ($r^2 = 0.226$). The cytochrome P450 gene (*CYP76A2*) on Chr10 showed a high value of ILD with an uncharacterized protein on Chr6.1 ($r^2 = 0.28$) as well as the intergenic region between the transcription factors *SPL1* and *DOF1.4*, both of which are responsible for plant growth and development, on Chr6.2 ($r^2 = 0.27$), perhaps indicating that this region on Chr6 may influence the regulation of the *CYP76A2* on Chr10 in some way. Furthermore, the highest linkage for Chr4-Chr10 ($r^2 = 0.274$), was observed for a SNP just upstream of *LOG3* on Chr4 and an uncharacterized protein on Chr10. Thus, the identified resistance alleles within these five highly differentiated regions show signs of linkage and perhaps evidence of epistatic selection.

Local regions of strong long-distance linkage disequilibrium and ILD within species might be aided by demographic processes like population structure (56, 57), genetic drift (58), or could be due to evolutionary processes like selection (28, 59). Furthermore, epistatic interactions among loci wherein adaptive alleles at two independent loci will be inherited together can generate linkage disequilibrium (28, 59–61). Given that our sampling design included multiple resistant and susceptible populations from varied locations, low population differentiation among populations, and evidence of recent migration between them (Supplementary Figure S4), it is unlikely that the observed ILD is due entirely to demographic processes. Moreover, we observed the strongest ILD between regions under selection harboring resistance associated genes, indicating the potential role of selection in maintaining the observed ILD. Thus, our finding suggests that the highly differentiated regions under selection containing candidate loci for glyphosate resistance are linked, indicating the potential role of epistatic selection in maintaining resistance.

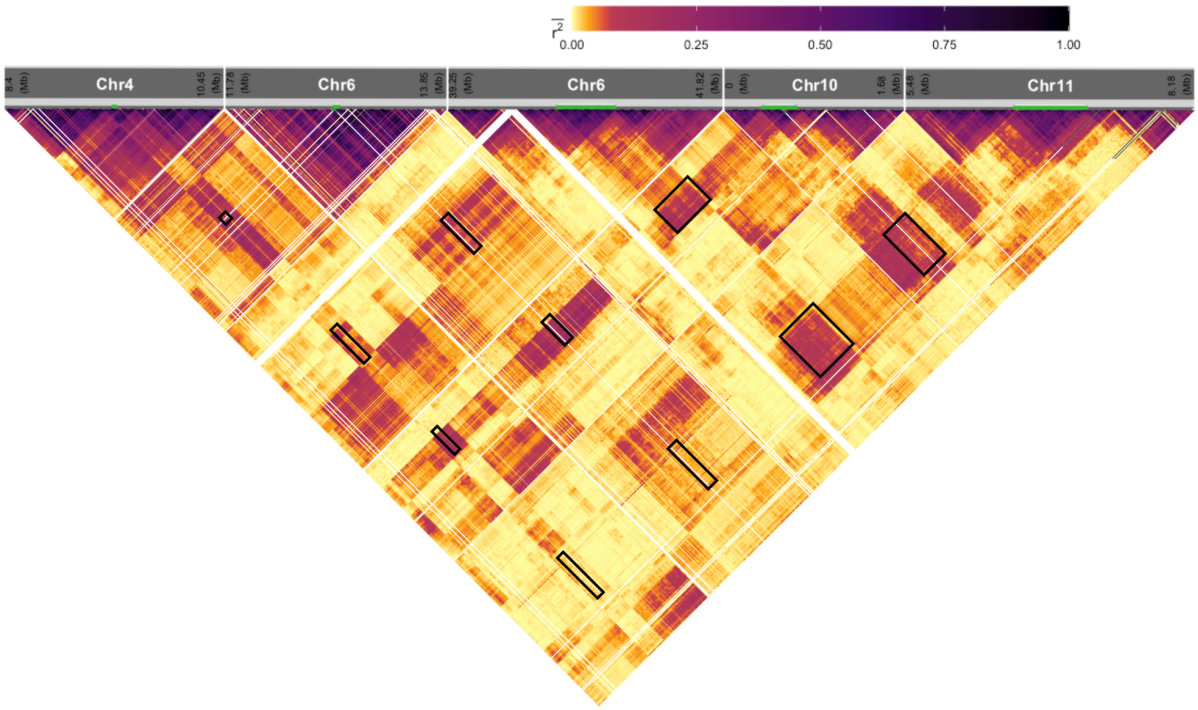


Figure 4 Long-distance linkage disequilibrium and ILD among the five highly differentiated (G_{ST} > 0.35) regions under selection associated with glyphosate resistance. The five intervals displayed islands of increased linkage disequilibrium as estimated by r^2 for SNPs separated by at least 1 kb in and between broad regions under selection. The white lines represent absence of SNPs (missing data) whereas the black boxes represent linkage between the five selection intervals. r^2 values are averaged over two-dimensional bins of 10 x 10 kb.

Signs of selection on potential cost loci

Our scan of regions associated with herbicide resistance, paired with a transcriptome survey, identified potential alleles with strong functional connections to the previously identified fitness cost in resistant *I. purpurea*. We found alternate alleles close to fixation in each population type within the 560kb highly differentiated region on chromosome 6 (40.25Mb - 40.81Mb; mean G_{ST} = 0.73, Figure 5a). This region contained the nuclear fission defective 6 (*NFD6*) and NAC transcription factor 25 (*NAC25*) genes, both of which function in seed development. *NAC25,* a gene that is required for normal seed development and morphology (Kunieda 2008), exhibited two missense variants in the resistant individuals (mutant allele resistant frequency = 0.91, susceptible frequency = 0.23). Additionally, *NFD6*, a protein required for nuclear fusion in the embryo sac during the production of the female gametophyte (62), contained six missense variants in the resistant individuals (mutant allele resistant frequency = 0.88, susceptible frequency = 0.21). The resistant haplotype of this gene also contained 10 SNPs in the promoter region which could

potentially alter its expression. Indeed, we found this protein to be downregulated in the presence of the herbicide, with a log-fold change of -3.52 in resistant individuals as compared to the susceptible individuals (Figure 5b). Thus, our data suggest that these genes may be responsible for the lower and abnormal germination leading to the observed fitness cost in this species (23).

Interestingly, this highly differentiated region containing these seed development genes is strongly linked to other regions under selection that harbor resistance alleles (99% ILD value = 0.22; Figure 4), indicating the potential role of linkage disequilibrium in maintaining the cost. More specifically, *NFD6* is within 83 kb from, and thus physically linked to (r^2 = 0.70), the potential regulatory region on Chr6.2 that exhibits interchromosomal long distance linkage disequilibrium with the CYP76A2 gene on Chr10 (ILD = 0.27). Further, the NAC25 gene is found in close proximity to serine/threonine kinases on Chr6.2 (i.e., 82 kb away), indicating another potential gene involved in the cost phenotype is in physical linkage (r^2 = 0.67) with potential NTSR loci.

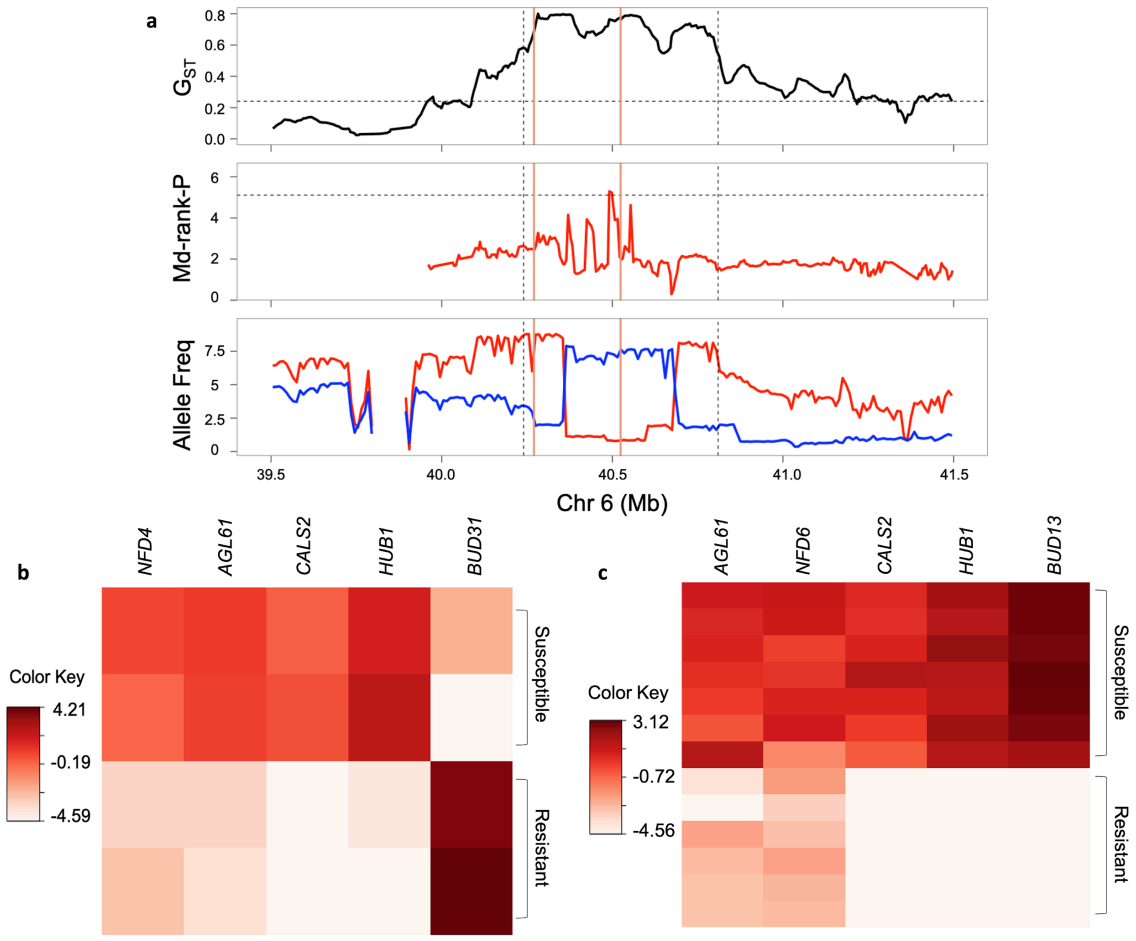


Figure 5 Loci associated with cost of glyphosate resistance identified by the (a) whole-genome selection-scan and differential expression analysis in the (b) absence and (c) presence of herbicide. Top panel of (a) represents G_{ST} between the resistant and the susceptible populations, mid panel is the Md-rank-*P* value, and the lower panel represents the allele frequency. Salmon vertical lines represent NFD6 and NAC25, in that order. Red and blue represent resistant and susceptible populations, respectively. Black horizontal dotted lines represent 95 percentile values while vertical lines represent regions with G_{ST} above 0.6. The differentially expressed cost genes shown here were chosen based on their functional annotation and had FDR < 0.005 and P-value < 0.00005. Color key represents log2 fold-change values.

In addition to the potential cost loci identified from the WGS screen, our gene expression analyses in the absence of herbicide (*i.e.* the environment in which fitness costs are assessed) comparing individuals from resistant and susceptible identified five differentially expressed gene that play a role in fertilization and seed maturation and are thus potentially related to the cost (Supplementary TableS6). Of special interest, the bud-site selection protein 31 (*BUD31*) was found to be highly upregulated (logFC = 7.22) in resistant plants in the absence of herbicide, whereas its homologue *BUD13* was highly significantly downregulated in resistant individuals in the presence of herbicide (logFC = -11.39). *BUD13* is involved in pre-mRNA splicing in embryos and is critical for early embryo development (63).

In the control environment, two genes downregulated in the resistant individuals -- *NFD4* (logFC = -3.61) and Agamous-like MADS-box protein AGL61 (*AGL61*; logFC = -4.98) -- are involved in megagametogenesis. The *NFD4* gene, like *NFD6*, is responsible for ovule polar nuclei fusion during female karyogamy (62), whereas *AGL61* is required for the central cell development and differentiation (64). A loss of function mutation in *AGL61* has been shown to cause abnormal morphology and over 50% seed abortion upon fertilization in Arabidopsis (64). We also identified a callose synthase 2 (*CALS2*) to be strongly downregulated in resistant individuals (logFC = -8.72); another member of the callose synthase family (*CALS5*) has been shown to be responsible for pollen viability (65). Finally, we also found that E3 ubiquitin-protein ligase BRE1 (*HUB1*), a protein involved in seed germination, was strongly downregulated among the resistant individuals (logFC = -8.27). *HUB1* has been shown to control chromatin remodeling during seed development and leads to alterations in seed dormancy (66).

Interestingly, three of these five candidate genes (*AGL61*, *CALS2*, *HUB1*), and homologues of other two (*BUD13* and *NFD6)* were also significantly downregulated in the resistant populations in the presence of herbicide (Supplementary TableS6). These candidate cost genes are all essential for plant reproduction and are highly downregulated (except *BUD31*) in the resistant population in both the absence and presence of the herbicide, and thus could potentially explain the phenotypic costs of glyphosate resistance in *I. purpurea* seen by Van Etten and colleagues (23).

Discussion

While there is an increasing appreciation for the role of nontarget site mechanisms underlying herbicide resistance in agricultural weeds (12, 67–69), there are strikingly few comprehensive whole genome assays of resistant weeds suggesting that the entirety of the NTSR response is rarely captured. Our study using a sequenced and assembled genome, whole genome resequencing of natural populations, and a gene expression survey offers a unique opportunity to identify loci associated with NTSR and to further investigate the evolutionary forces that underlie the maintenance and expected evolutionary trajectory of resistance alleles in natural populations.

Our results show detoxification underlies resistance in *I. purpurea*. Detoxification is hypothesized to involve three steps -- uptake of the herbicide by phosphate transporters, chemical modification (*i.e.* the addition of an OH and sugar group by cytochrome P450s and glycosyltransferases, respectively), and transport to vacuoles by ABC transporters and other sugar transporters where the molecule is stored and/or inactivated (12, 18). We found evidence of selection on all genes involved in this pathway. We also found evidence of selection (and in some cases, differential expression) of genes involved in plant signalling and environmental stress (i.e., serine/threonine kinases, ERFs, E3-Ubiquitin Ligases, FRSs, LOG3, MDAR4, GT-3B) as well as genes involved in the shikimate acid pathway (DHD/SHD, *TYRAAT2*). Our results thus expand what we currently know about the detoxification NTSR mechanism in this species to include plant signalling and stress responses, both of which are either hypothesized (7) or shown to be involved in herbicide resistance (70–73). While we do not currently have functional genomics resources for this species, our study using a cytochrome P450 inhibitor verifies that resistant *I. purpurea* individuals have the ability to detoxify the herbicide. The next step in understanding resistance in *I. purpurea* involves determining the contribution of each of the candidate loci under selection (and/or showing differential regulation) to both resistance and its associated cost. With future development of genome editing protocols for *I. purpurea*, we will be able to experimentally test the function of loci hypothesized to be contributing to herbicide resistance.

Due to the involvement of multiple genes involved in the herbicide detoxification pathway, and evidence for selection on regions of the genome found on separate chromosomes, we hypothesized that multiple loci would show evidence of ILD, perhaps indicating epistatic selection between alleles or co-inheritance. Our results support this hypothesis. Foremost, in contrast to low background ILD, long-distance linkage disequilibrium and ILD were high among intervals under selection, and consistently differentiated between the resistant and susceptible types across multiple populations. The strongest linkage was observed between putative resistance genes that exhibited signs of selection. This linkage could quickly become very steep in the presence of epistatic interactions among loci (74), as

would be the case if genes underlying NTSR worked in concert to produce the resistance phenotype. Indeed, we found high ILD values between regulatory regions and resistance alleles, and between intervals harboring genes involved in the same molecular pathways (e.g. detoxification, and stress signaling and response).

Although linkage should become decoupled over time due to recombination and gene flow, the ongoing selection for herbicide resistance could slow down this decoupling between these functionally interacting genes (75). Even given gene flow between these populations (24), co-adaptation and epistasis could lead to the fixation of the resistance alleles given strong selection (76) or weaker recombination rate (77). Thus, long-distance linkage disequilibrium and ILD aided with epistatic selection could act to maintain resistance through generations in natural populations.

One evolutionary force that should counteract the continued evolution of resistance is the potential for fitness costs of resistance, either due to the pleiotropic effects of resistance alleles themselves or due to negative fitness effects of loci that are linked to resistance loci. While costs are central to theories of resistance evolution (78–81), there are currently no examples, to our knowledge, in which the loci underlying fitness costs of nontarget site resistance have been identified. Our results suggest candidate loci associated with the previously identified cost of glyphosate resistance. Specifically, we found a highly differentiated region on Chr6 that exhibited alternate alleles in resistant and susceptible populations, and found this region to contain loci required for normal seed development and maturation (*NAC25, NFD6*). One of these genes, *NFD6*, was differentially regulated in the resistant individuals, further supporting its role in the low seed quality, and thus fitness cost, that we have previously described (23).

Additionally, our results strongly suggest genetic hitchhiking may act to maintain the cost in this species. Both *NAC25* and *NFD6* are physically linked on chromosome 6 to the regions under selection containing serine-threonine kinase genes and a regulatory region that is itself exhibiting ILD to the *CYP76A2* gene on chromosome 10. Although recombination should decouple cost alleles that are physically linked to resistance alleles, these loci would not completely decouple if the recombination rate (c) is much lower than the selection coefficient (s) (i.e., $c \ll s$, (82). The requirement that $c \ll s$ is not improbable given the close proximity of cost and resistance loci (< 85kb) and the strong ongoing selection for herbicide resistance. Furthermore, if the ratio $c/s < 10^{-4}$, the hitchhiking would almost be complete and the cost alleles could become fixed in the populations (83). Alternatively, it is possible that new compensatory mutations arising in the population could increase in frequency over time due to selection, decoupling the cost and resistance alleles and thus reducing fitness cost associated with glyphosate resistance (84, 85).

Overall, our work identified the potential genetic basis of NTSR glyphosate resistance in *I. purpurea* -- our whole genome and transcriptome assays strongly support the role of detoxification conferring herbicide resistance in this species, and we additionally identified a role for plant sensing and stress along with components of the shikimate acid pathway. Interestingly, we show that NTSR glyphosate resistance in *I. purpurea* is conferred by multiple loci which are maintained through generations via ILD. We also provide strong evidence to support the idea that fitness costs may be due to loci in strong linkage with resistance loci. Our work highlights the importance of multi-level, multi-population study in identifying the genetic mechanisms underlying polygenic defense traits, and for understanding the complex genetic-interplay between defense and cost.

Methods

Genome sequencing, assembly, and annotation: We used an *I. purpurea* line originally sampled from an agricultural field in Orange County, NC, in 1989 by M. Rausher (*i.e.*, prior to the widespread use of glyphosate) and selfed for >18 generations in the lab for genome sequencing (seeds of this line 'Fred/C' are available upon request). High molecular weight DNA was isolated from flash-frozen leaf tissue using a modified large-volume CTAB protocol (86) and sequenced on a PacBio Sequel at the University of Georgia. Raw PacBio subreads from 9 cells of Sequel chemistry were error-corrected with Canu (v1.7.1) (87) with default parameters for raw PacBio reads (--pacbio-raw). The corrected and trimmed reads from Canu were assembled with Flye (v2.4-release) (88) and anchored onto pseudomolecules by nearly 81 million read pairs of Phase Genomics Hi-C (Seattle, WA) of leaf tissue using Sau3AI cutsites. Within-genome and across-genome synteny was visualized using the CoGE SynMap platform (89), with DAGChainer options "-D20 -A 5", as well as with jcvi with default parameters (https://github.com/tanghaibao/jcvi). *Ipomoea purpurea* pseudomolecules were numbered and oriented according to chromosome synteny against *Ipomoea nil* pseudomolecules (Supplemental Figure S2).

Raw 50nt single-end RNA-seq reads were aligned using STAR (v.2.7.0) (90) with default single-pass parameters. Repetitive elements were first annotated with RepeatModeler (v1.0.11). Long Terminal Repeat (LTR) retrotransposons were annotated with LTRharvest (v1.6.1) with options -similar 85 -mindistltr 1000 -maxdistltr 15000 -mintsd 5 -maxtsd 20". RepeatModeler annotations were combined with all Viridiplantae repeats from Repbase and used as a species-specific repeat database built using RepeatModeler with default options.

Genome annotation was performed using a diverse set of evidence. First, a set of 12 RNA-seq libraries from leaf tissue was aligned with STAR (v2.7.0), and transcripts assembled with Stringtie (v2.1.3). MAKER2 (91) was initially run with evidence from the RNA-seq alignments, as well as peptides from *I. trifida, I. triloba,* and *I. nil*. The resulting gene set was used to train SNAP (v2013-11-29). AUGUSTUS (v3.3.2) was trained with evidence from BUSCO (v4.1.0) against the eudicot odb10 set. with default options. MAKER2

was re-run with the *ab initio* SNAP and AUGUSTUS training sets, in addition to the homologous protein and RNA-seq evidence, to build a final gene annotation set.

Sampling and sequencing: We selected eight populations to investigate the genetic basis of glyphosate resistance and its cost following (20)-- 4 low resistance, from here on referred to as the susceptible population (S: <20% population survival at 1X the field dose of RoundUp) and 4 high resistance populations (R: >70% population survival at 1X the field dose of RoundUp), from here on referred to as the resistant populations (Supplementary TableS1). Seeds from 10 maternal lines per population were germinated, except for one susceptible population (RB), wherein 9 maternal lines were used. We extracted DNA from leaf tissue using the Qiagen Plant DNeasy kit. 150 paired-end sequencing was performed using Illumina HiSeq4000 and NovaSeq6000 using three and two lanes, respectively. We sequenced two populations at high coverage (at least 25X) and the remaining six populations at low coverage (10X). Two populations (WG, resistant and RB, susceptible) were run on one lane of HiSeq6000 and NovaSeq6000 each whereas the other lane had the remaining six populations. This yielded a total of 3,300,397,148,700 bases with average coverage of 28.84X for WG and RB. Coverage of the other six populations has an average of 14.66X.

Variant calling: We aligned the reads to our draft genome using BWA mem v0.7.15 (92) with parameter -M. Since the same sample was sequenced using multiple platforms (HiSeq and NovaSeq), the alignment files were merged and duplicate reads were marked using the MarkDuplicate tool of Picard v2.8.1 (http://broadinstitute.github.io/picard). Next, we prepared a database of true known variants, required for base recalibration. This database was created using data from the top eighteen individuals with the highest read counts, upon which variant call was performed using the HaplotypeCaller tool of GATK v4.1 (93). Low confidence variants were filtered out using the VariantFiltration tool of GATK v4.1 (93) (15 < DP < 60; ReadPosRankSum < -8.0; QD < 2.0; FS > 60.0; SOR > 3.0; MQ < 40.0; MQRankSum < -12.5) and only the high confidence variants were used in the dataset. This was used to recalibrate base qualities using GATK v4.1 tools BaseRecalibrator and ApplyBQSR (93). Variants were called individually on all the individuals using the HaplotypeCaller tool of GATK v4.1 (93) using parameters -ERC GVCF --min-pruning 1 --min-dangling-branch-length 1. The variants from each individual were combined to one variant file (a raw cohort variant file) using the tools GenomicsDBImport, GenotypeGVCFs, and GatherVcfs (93). Next, multiple rounds of filtration were performed on this variant dataset to filter out potential false positives. First, using the GATK v4.1 tools VariantFiltration and SelectVariants we filtered the variants using the parameters QD<1.5, DP<10 and DP>5000, FS>60, SOP>5, MQ<40 and MQ>100, MQRankSum< -6 and MQRankSum>6, and ReadPosRankSum< -4 and ReadPosRankSum> 6 (93). For the next round of filtration, we removed variants that had genotype depth more than twice the

average and heterozygosity more than 0.8 using the geno-depth and het packages from VCFtools v0.1.15 (94), and a custom R script (github link). In the third round of filtration, we filtered variants that had quality above 20, had no missing information, a minor allele frequency of 0.2, and a minimum mean depth of 10 (vcftools --minQ 20 --max-missing 1.0 --maf 0.2 --min-meanDP 10) (94). Finally, we filtered using BCFtools (v1.7) (95) to keep only bi-allelic SNPs (bcftools view -m2 -M2 -v snps). This gave us a total of 4,909,800 high confidence SNPs. These SNPs were used for downstream analyses.

We performed a PCA analysis using the allele frequencies of all the SNPs to investigate the population structure using the package bigsnpr v.1.4.4 (96) in R, and found that the populations did not segregate into two separate genetic clusters (Supplementary Figure S4a-b). Further, we repeated this analysis for SNPs from regions under selection (see below) to test whether we observe the same population structure patterns. We observed that these separated into distinct resistant and susceptible groups, with the exception of a resistant population, BI, which clustered between the susceptible and other resistant populations (Supplementary Figure S4c). Thus for the purposes of this study, we dropped the BI population from further analysis.

Selection analysis: We split the high confidence variant dataset obtained into 'resistant' and 'susceptible' variant dataset using vcf-subset of VCFtools v0.1.15 (94). The 'resistant' and 'susceptible' variant datasets contained 30 and 39 individuals, respectively (Table S1). We then used these datasets to calculate diversity and selection statistics G_{ST} (97), pi, Tajima's D (98), Fu and Way's H (99) using a custom script from (100) in a 300SNP window, for both the dataset. Furthermore, to detect hard sweep we phased the variants using beagle version 5.1 (101) which was then used to calculate the haplotype homozygosity statistic (H12, a measure of haplotype homozygosity that detects both hard and soft sweeps) using the scripts provided (102). For regions above 95 percentile G_{ST}, we calculated a composite rank based statistic (Md-rank-*P*) which was computed as the Mahalanobis distance on the negative log10 transformation of raw statistics into rank P-values (103). This Md-rank-P was calculated using pi, Tajima's D, Fu and Way's H, and H12. To identify potential regions of selection we chose bins with greater than 95 percentile Md-rank-*P*.

Linkage analysis: We calculated linkage disequilibrium (r^2) at three different levels. First, to estimate the genome-wide linkage disequilibrium with physical distance and its decay, we calculated the pairwise r^2 for 30 randomly selected SNPs within non-overlapping 100kb windows using Plink version 1.9 (104) (--r2 --ld-window-r2 0.01 --ld-window 999999 --ld-window-kb 65000). For all SNPs within this subset, a mean-SNP pairwise r^2 value physically within 100kb distance was calculated. We further used this to calculate the mean r^2 for each chromosome and the mean r^2 within 10kb windows. Second, to estimate the background genome-wide long-distance (and interchromosomal) linkage disequilibrium

(ILD), we calculated r^2 values for 5842 SNPs separated by at least 100kb using VCFtools v0.1.15 (94) (--thin 100000 --interchrom-hap-r2). Third, we estimated the r^2 for SNPs separated by at least 1kb in and between broad regions (1Mb to 1.5Mb) around the five focused regions (with $G_{ST} > 0.35$) under selection using VCFtools v0.1.15 (94).

RNA-Seq: To identify transcripts associated with glyphosate resistance and its potential cost, we sequenced transcriptomes of 17 individuals belonging to four different treatments; resistant control (Rc), susceptible control (Sc), resistant herbicide sprayed (Rh), and susceptible herbicide sprayed (Sh). Each treatment had multiple individuals (Rc-2, Sc-2, Rh-6, Sh-7; Supplementary TableS2). The seeds were grown in a controlled environment (growth chamber) to reduce variation due to environmental differences. 20 days after planting, we sprayed glyphosate (concentration of 1.52 kg ai/ha) on the Rh and Sh treatment plants and collected the second and fourth leaf for RNA extractions 8 hours post-spray. These were flash frozen using liquid nitrogen and stored at -80°C. We extracted RNA using Qiagen RNeasy Plant mini kit with the optional DNase digestion step. This was then sequenced using Illumina NovaSeq 6000 at 150bp paired-end sequencing. A total of 132,551,535,000 bp were obtained.

Differential gene expression -- We processed the raw reads obtained to remove adapters using cutadapt v1.18 (105) and then mapped them to the de-novo assembled genome (--sjdbOverhang 149 --outSAMtype BAM SortedByCoordinate Unsorted) using STAR v2.7.5 (90). Next, using HTSeq v0.11.1 (106), we counted read counts for each gene. These read counts were then used to filter out lowly expressed transcripts using the Bioconductor package edgeR version 3.18.1 (107) such that transcripts were retained only if they had greater than 0.5 counts-per-million in at least two samples (Rc vs Sc) and four samples (Rh vs Sh). The libraries were then normalized in edgeR (using the trimmed mean of M-values method) followed by differential gene expression analysis using the classic pairwise comparison of edgeR version 3.18.1. We extracted the significance of differentially expressed transcripts (DETs) with FDR <= 0.05. This was done for two contrasts, Rh vs Sh (total sample size = 13; Rh = 6, Sh = 7) and Rc vs Sc (total sample size 4; Rc = 2, Sc = 2). The first contrast informs us of the genes that are regulated in response to the herbicide, and how this gene regulation differs between the resistant and the susceptible populations, whereas the latter informs us of the baseline expression difference due to glyphosate resistance between the two populations.

Malathion Experiment: On May 15th, 2019, we planted a total of 180 replicate seeds from multiple resistant and susceptible populations (Supplementary TableS3) in Cone-Tainers (Stewe and Sons). These were allowed to grow for 30 days, after which we subjected them to one of the four treatment environments--malathion (7.81 ml/L according to manufacturer's recommendations), glyphosate (3.4 kg ai/ha), glyphosate and malathion,

and a control. Twenty-five days post treatment spray, we recorded death as a metric trait (dead/almost dead or green and healthy), and harvested the plants. These were dried for 3 days at 70C and weighed for an estimate of dry above ground-biomass.

Using this data, we assessed whether biomass were significantly altered by the different treatments. First, we normalized the above-ground biomass using the transformTukey function from Rcompanion v.2.0.0 (108). We then used a generalized linear model (lm function;(109) with as normalized biomass as the dependent variable and population Type (R/S) and treatment as the independent variables. We assessed the significance of the variables using the Anova function of the car package v.3.0.10 (110), and performed a pairwise comparison between groups using the lsmeans function from package lsmeans v2.30.0 (111), adjusted for multiple tests using tukey correction. Using the same general model, we also compared whether biomass was significantly different between treatments for each population type. To control for the differences in the plant size we standardized the biomass of the individuals by the average biomass of the respective maternal line in the control treatment, and then normalized it as above.

Data Availability

The *I. purpurea* genome assembly and annotation are available in the CoGe platform - https://genomevolution.org/coge/GenomeInfo.pl?gid=58735 .

Acknowledgements
We thank M. Rausher for supplying germplasm used in the *I. purpurea* genome sequencing, and T. Newsum, S. Paranjape, and K. Johnson for growth room phenotyping, and M. Palmer and MBGNA for growth room support. We likewise thank J. Opp in the University of Michigan sequencing core for sequencing support, Amanda Cummings for high molecular weight DNA preparation, and the Georgia Genomics and Bioinformatics Core, which provided the PacBio Sequel sequencing service. We also thank E.B. Josephs and R.L. Rogers for providing feedback on the manuscript. Funding for this work was provided by the University of Michigan and USDA NIFA awards 24892 & 28497 to RSB, NSF DEB 1442199 and IOS 1444567 to J.L.M., and NSF IOS 1611853 to A.H.

Footnotes

[1]S.G. and A.H. contributed equally to this work.

[2]Author contributions: R.S.B., and S.G. designed the project; R.S.B., M.L.V.E., and S.G. designed the experiments; A.H. and J.L.M. assembled and annotated the genome; A.S. performed the growth room experiment; M.L.V.E. performed the WGS experiment, S.G.

analyzed whole genome resequencing, RNAseq data, and performed linkage analysis; and R.S.B., S.G., A.H., and J.L.M. wrote the paper.

The authors declare no competing interest.

References

1. N. Liu, Insecticide resistance in mosquitoes: impact, mechanisms, and research directions. *Annu. Rev. Entomol.* **60**, 537–559 (2015).

2. N. J. Hawkins, C. Bass, A. Dixon, P. Neve, The evolutionary origins of pesticide resistance. *Biol. Rev. Camb. Philos. Soc.* (2018) https:/doi.org/10.1111/brv.12440.

3. S. B. Powles, Q. Yu, "Evolution in Action: Plants Resistant to Herbicides" in Annual Review of Plant Biology., S. Merchant, W. R. Briggs, D. Ort, Eds. (2010), pp. 317–347.

4. J. Mithila, A. S. Godar, Understanding Genetics of Herbicide Resistance in Weeds: Implications for Weed Management. *Advances in Crop Science and Technology* **1**, 1–3 (2013).

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

6. B. P. Murphy, P. J. Tranel, Target-Site Mutations Conferring Herbicide Resistance. *Plants* **8** (2019).

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

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

9. H. J. Beckie, Herbicide Resistance in Plants. *Plants* **9** (2020).

10. C. Délye, M. Jasieniuk, V. Le Corre, Deciphering the evolution of herbicide resistance in weeds. *Trends Genet.* **29**, 649–658 (2013).

11. T. A. Gaines, E. L. Patterson, P. Neve, Molecular mechanisms of adaptive evolution revealed by global selection for glyphosate resistance. *New Phytol.* **223**, 1770–1775 (2019).

12. T. A. Gaines, *et al.*, Mechanisms of evolved herbicide resistance. *J. Biol. Chem.* **295**, 10307–10330 (2020).

13. R. Busi, P. Neve, S. Powles, Evolved polygenic herbicide resistance in Lolium rigidum by low-dose herbicide selection within standing genetic variation. *Evol. Appl.* **6**, 231–242 (2013).

14. L. Scarabel, F. Pernin, C. Délye, Occurrence, genetic control and evolution of non-target-site based resistance to herbicides inhibiting acetolactate synthase (ALS) in the dicot weed Papaver rhoeas. *Plant Sci.* **238**, 158–169 (2015).

15. J. M. Kreiner, J. R. Stinchcombe, S. I. Wright, Population Genomics of Herbicide Resistance: Adaptation via Evolutionary Rescue. *Annu. Rev. Plant Biol.* **69**, 611–635 (2018).

16. I. Cummins, *et al.*, Key role for a glutathione transferase in multiple-herbicide resistance in grass weeds. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 5812–5817 (2013).

17. B. A. Pandian, R. Sathishraj, P. V. V. Prasad, M. Jugulam, A single gene inherited trait confers metabolic resistance to chlorsulfuron in grain sorghum (Sorghum bicolor). *Planta* **253**, 48 (2021).

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

19. J. M. Kreiner, P. J. Tranel, D. Weigel, J. R. Stinchcombe, S. I. Wright, The genetic architecture and genomic context of glyphosate resistance. *Cold Spring Harbor Laboratory*, 2020.08.19.257972 (2020).

20. M. Van Etten, K. M. Lee, S.-M. Chang, R. S. Baucom, Parallel and nonparallel genomic responses contribute to herbicide resistance in Ipomoea purpurea, a common agricultural weed. *PLoS Genet.* **16**, e1008593 (2020).

21. A. Kuester, S.-M. Chang, R. S. Baucom, The geographic mosaic of herbicide resistance evolution in the common morning glory, Ipomoea purpurea: Evidence for resistance hotspots and low genetic differentiation across the landscape. *Evol. Appl.* **8**, 821–833 (2015).

22. A. Kuester, A. Wilson, S.-M. Chang, R. S. Baucom, A resurrection experiment finds evidence of both reduced genetic diversity and potential adaptive evolution in the agricultural weed Ipomoea purpurea. *Mol. Ecol.* **25**, 4508–4520 (2016).

23. M. L. Van Etten, A. Kuester, S.-M. Chang, R. S. Baucom, Fitness costs of herbicide resistance across natural populations of the common morning glory, Ipomoea purpurea. *Evolution* **70**, 2199–2210 (2016).

24. D. F. Alvarado-Serrano, M. L. Van Etten, S.-M. Chang, R. S. Baucom, The relative contribution of natural landscapes and human-mediated factors on the connectivity of a noxious invasive weed. *Heredity* **122**, 29–40 (2019).

25. D. B. Lowry, *et al.*, Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Mol. Ecol. Resour.* **17**, 142–152 (2017).

26. P. M. Petkov, *et al.*, Evidence of a large-scale functional organization of mammalian chromosomes. *PLoS Genet.* **1**, e33 (2005).

27. Q. Long, *et al.*, Massive genomic variation and strong selection in Arabidopsis

thaliana lines from Sweden. *Nat. Genet.* **45**, 884–890 (2013).

28. K. Hench, M. Vargas, M. P. Höppner, W. O. McMillan, O. Puebla, Inter-chromosomal coupling between vision and pigmentation genes during genomic divergence. *Nature Ecology & Evolution* **3**, 657–667 (2019).

29. B. Wallace, On Coadaptation in Drosophila. *Am. Nat.* **87**, 343–358 (1953).

30. T. Dobzhansky, *Genetics of the Evolutionary Process* (Columbia University Press, 1971).

31. D. Schluter, *The Ecology of Adaptive Radiation* (OUP Oxford, 2000).

32. A. Hoshino, *et al.*, Genome sequence and analysis of the Japanese morning glory Ipomoea nil. *Nat. Commun.* **7**, 13295 (2016).

33. D. G. Hardie, PLANT PROTEIN SERINE/THREONINE KINASES: Classification and Functions. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**, 97–131 (1999).

34. J.-H. Lee, W. T. Kim, Regulation of abiotic stress signal transduction by E3 ubiquitin ligases in Arabidopsis. *Mol. Cells* **31**, 201–208 (2011).

35. K. Thirugnanasambantham, *et al.*, Role of Ethylene Response Transcription Factor (ERF) and Its Regulation in Response to Stress Encountered by Plants. *Plant Mol. Biol. Rep.* **33**, 347–357 (2015).

36. W. Liu, *et al.*, The ethylene response factor AtERF4 negatively regulates the iron deficiency response in Arabidopsis thaliana. *PLoS One* **12**, e0186580 (2017).

37. L. Ma, G. Li, FAR1-RELATED SEQUENCE (FRS) and FRS-RELATED FACTOR (FRF) Family Proteins in Arabidopsis Growth and Development. *Front. Plant Sci.* **9**, 692 (2018).

38. J. H. Kim, W. T. Kim, The Arabidopsis RING E3 ubiquitin ligase AtAIRP3/LOG2 participates in positive regulation of high-salt and drought stress responses. *Plant Physiol.* **162**, 1733–1749 (2013).

39. P. J. Eastmond, MONODEHYROASCORBATE REDUCTASE4 is required for seed storage oil hydrolysis and postgerminative growth in Arabidopsis. *Plant Cell* **19**, 1376–1387 (2007).

40. H. C. Park, *et al.*, Pathogen- and NaCl-induced expression of the SCaM-4 promoter is mediated in part by a GT-1 box that interacts with a GT-1-like transcription factor. *Plant Physiol.* **135**, 2150–2161 (2004).

41. L. Ding, *et al.*, Functional analysis of the essential bifunctional tobacco enzyme 3-dehydroquinate dehydratase/shikimate dehydrogenase in transgenic tobacco plants. *J. Exp. Bot.* **58**, 2053–2067 (2007).

42. P. Rippert, M. Matringe, Purification and kinetic analysis of the two recombinant arogenate dehydrogenase isoforms of Arabidopsis thaliana. *Eur. J. Biochem.* **269**, 4753–4761 (2002).

43. H. C. Steinrücken, N. Amrhein, The herbicide glyphosate is a potent inhibitor of 5-enolpyruvyl-shikimic acid-3-phosphate synthase. *Biochem. Biophys. Res. Commun.* **94**, 1207–1212 (1980).

44. E. Walper, C. Weiste, M. J. Mueller, M. Hamberg, W. Dröge-Laser, Screen Identifying Arabidopsis Transcription Factors Involved in the Response to 9-Lipoxygenase-Derived Oxylipins. *PLoS One* **11**, e0153216 (2016).

45. Z. Xie, *et al.*, The AP2/ERF Transcription Factor TINY Modulates Brassinosteroid-Regulated Plant Growth and Drought Responses in Arabidopsis. *Plant Cell* **31**, 1788–1806 (2019).

46. A. Takemiya, *et al.*, Phosphorylation of BLUS1 kinase by phototropins is a primary step in stomatal opening. *Nat. Commun.* **4**, 2094 (2013).

47. S. Liu, *et al.*, Negative feedback regulation of ABA biosynthesis in peanut (Arachis hypogaea): a transcription factor complex inhibits AhNCED1 expression during water stress. *Sci. Rep.* **6**, 37943 (2016).

48. Z. Lai, K. Vinod, Z. Zheng, B. Fan, Z. Chen, Roles of Arabidopsis WRKY3 and WRKY4 transcription factors in plant responses to pathogens. *BMC Plant Biol.* **8**, 68 (2008).

49. J. Jiang, *et al.*, WRKY transcription factors in plant responses to stresses. *J. Integr. Plant Biol.* **59**, 86–101 (2017).

50. X.-Y. Zhao, *et al.*, The MdWRKY31 transcription factor binds to the MdRAV1 promoter to mediate ABA sensitivity. *Horticulture Research* **6**, 66 (2019).

51. C. Piasecki, *et al.*, Transcriptomic Analysis Identifies New Non-Target Site Glyphosate-Resistance Genes in Conyza bonariensis. *Plants* **8** (2019).

52. J. T. Christopher, C. Preston, S. B. Powles, Malathion Antagonizes Metabolism-Based Chlorsulfuron Resistance in Lolium rigidum. *Pestic. Biochem. Physiol.* **49**, 172–182 (1994).

53. C. Preston, F. J. Tardif, J. T. Christopher, Multiple Resistance to Dissimilar Herbicide Chemistries in a Biotype ofLolium rigidumDue to Enhanced Activity of Several Herbicide Degrading Enzymes. *Pestic. Biochem. Physiol.* (1996).

54. Q. Yu, S. Powles, Metabolism-Based Herbicide Resistance and Cross-Resistance in Crop Weeds: A Threat to Herbicide Sustainability and Global Crop Production. *Plant Physiol.* **166**, 1106–1118 (2014).

55. M. Yanniccari, R. Gigón, A. Larsen, Cytochrome P450 Herbicide Metabolism as the Main Mechanism of Cross-Resistance to ACCase- and ALS-Inhibitors in Lolium spp. Populations From Argentina: A Molecular Approach in Characterization and Detection. *Front. Plant Sci.* **11**, 600301 (2020).

56. M. Nei, W. H. Li, Linkage disequilibrium in subdivided populations. *Genetics* **75**, 213–219 (1973).

57. J. F. Wilson, D. B. Goldstein, Consistent Long-Range Linkage Disequilibrium Generated by Admixture in a Bantu-Semitic Hybrid Population. *Am. J. Hum. Genet.* **67**, 926–935 (2000).

58. E. Schaper, A. Eriksson, M. Rafajlovic, S. Sagitov, B. Mehlig, Linkage disequilibrium under recurrent bottlenecks. *Genetics* **190**, 217–229 (2012).

59. P. A. Hohenlohe, S. Bassham, M. Currey, W. A. Cresko, Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 395–408 (2012).

60. M. Sohail, *et al.*, Negative selection in humans and fruit flies involves synergistic epistasis. *Science* **356**, 539–542 (2017).

61. P. Behrouzi, E. C. Wit, Detecting Epistatic Selection with Partially Observed Genotype Data Using Copula Graphical Models. *arXiv [stat.AP]* (2017).

62. M. F. Portereiko, *et al.*, NUCLEAR FUSION DEFECTIVE1 encodes the Arabidopsis RPL21M protein and is required for karyogamy during female gametophyte development and fertilization. *Plant Physiol.* **141**, 957–965 (2006).

63. F. Xiong, *et al.*, AtBUD13 affects pre-mRNA splicing and is essential for embryo development in Arabidopsis. *Plant J.* **98**, 714–726 (2019).

64. J. G. Steffen, I.-H. Kang, M. F. Portereiko, A. Lloyd, G. N. Drews, AGL61 interacts with AGL80 and is required for central cell development in Arabidopsis. *Plant Physiol.* **148**, 259–268 (2008).

65. X. Dong, Z. Hong, M. Sivaramakrishnan, M. Mahfouz, D. P. S. Verma, Callose synthase (CalS5) is required for exine formation during microgametogenesis and for pollen viability in Arabidopsis. *Plant J.* **42**, 315–328 (2005).

66. Y. Liu, *et al.*, Identification of the Arabidopsis REDUCED DORMANCY 2 gene uncovers a role for the polymerase associated factor 1 complex in seed dormancy. *PLoS One* **6**, e22241 (2011).

67. C. Délye, J. A. C. Gardin, K. Boucansaud, B. Chauvel, C. Petit, Non-target-site-based resistance should be the centre of attention for herbicide resistance research: Alopecurus myosuroides as an illustration: Why we need more research on NTSR.

Weed Res. **51**, 433–437 (2011).

68. H. Ghanizadeh, K. C. Harrington, Non-target Site Mechanisms of Resistance to Herbicides. *CRC Crit. Rev. Plant Sci.* **36**, 24–34 (2017).

69. M. Jugulam, C. Shyam, Non-Target-Site Resistance to Herbicides: Recent Developments. *Plants* **8** (2019).

70. D. E. M. Radwan, Salicylic acid induced alleviation of oxidative stress caused by clethodim in maize (Zea mays L.) leaves. *Pestic. Biochem. Physiol.* **102**, 182–188 (2012).

71. A. Duhoux, C. Délye, Reference genes to study herbicide stress response in Lolium sp.: up-regulation of P450 genes in plants resistant to acetolactate-synthase inhibitors. *PLoS One* **8**, e63576 (2013).

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

73. T. Vega, *et al.*, Stress response and detoxification mechanisms involved in non-target-site herbicide resistance in sunflower. *Crop Sci.* **60**, 1809–1822 (2020).

74. R. C. Lewontin, K.-I. Kojima, The Evolutionary Dynamics of Complex Polymorphisms. *Evolution* **14**, 458–472 (1960).

75. M. Nei, Modification of linkage intensity by natural selection. *Genetics* **57**, 625–641 (1967).

76. K. R. Takahasi, Evolution of coadaptation in a subdivided population. *Genetics* **176**, 501–511 (2007).

77. K. R. Takahasi, F. Tajima, Evolution of coadaptation in a two-locus epistatic system. *Evolution* **59**, 2324–2332 (2005).

78. E. L. Simms, M. D. Rausher, Costs and Benefits of Plant Resistance to Herbivory. *Am. Nat.* **130**, 570–581 (1987).

79. E. A. Stahl, G. Dwyer, R. Mauricio, M. Kreitman, J. Bergelson, Dynamics of disease resistance polymorphism at the Rpm1 locus of Arabidopsis. *Nature* **400**, 667–671 (1999).

80. R. S. Baucom, R. Mauricio, Fitness costs and benefits of novel herbicide tolerance in a noxious weed. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 13386–13390 (2004).

81. M. M. Vila-Aiub, P. Neve, S. B. Powles, Fitness costs associated with evolved herbicide resistance alleles in plants. *New Phytol.* **184**, 751–767 (2009).

82. W. Stephan, T. H. E. Wiehe, M. W. Lenz, The effect of strongly selected

substitutions on neutral polymorphism: Analytical results based on diffusion theory. *Theor. Popul. Biol.* **41**, 237–254 (1992).

83. J. C. Fay, C. I. Wu, Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405–1413 (2000).

84. T. Vogwill, M. Kojadinovic, R. C. MacLean, Epistasis between antibiotic resistance mutations and genetic background shape the fitness effect of resistance across species of Pseudomonas. *Proc. Biol. Sci.* **283** (2016).

85. T. Lenormand, N. Harmand, R. Gallet, Cost of resistance: an unreasonably expensive concept. *ReEco* **3**, 51–70 (2018).

86. J. J. Doyle, J. L. Doyle, Isolation ofplant DNA from fresh tissue. *Focus* **12**, 39–40 (1990).

87. S. Koren, *et al.*, Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).

88. M. Kolmogorov, J. Yuan, Y. Lin, P. Pevzner, Assembly of Long Error-Prone Reads Using Repeat Graphs. *bioRxiv*, 247148 (2018).

89. E. H. Lyons, *CoGe, a new kind of comparative genomics platform: Insights into the evolution of plant genomes* (University of California, Berkeley, 2008).

90. A. Dobin, *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

91. C. Holt, M. Yandell, MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011).

92. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

93. A. McKenna, *et al.*, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

94. P. Danecek, *et al.*, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

95. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).

96. F. Privé, H. Aschard, A. Ziyatdinov, M. G. B. Blum, Efficient analysis of large-scale genome-wide data with two R packages: bigstatsr and bigsnpr. *Bioinformatics* **34**, 2781–2787 (2018).

97. B. S. Weir, Genetic Data Analysis II Sinauer Associates. *Inc. , Sunderland, MA* (1996).

98. F. Tajima, Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).

99. J. C. Fay, C.-I. Wu, Sequence Divergence, Functional Constraint, and Selection in Protein Evolution. *Annu. Rev. Genomics Hum. Genet.* **4**, 213–235 (2003).

100. P. Baduel, B. Hunter, S. Yeola, K. Bomblies, Genetic basis and evolution of rapid cycling in railway populations of tetraploid Arabidopsis arenosa. *PLoS Genet.* **14**, e1007510 (2018).

101. S. R. Browning, B. L. Browning, Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).

102. N. R. Garud, P. W. Messer, E. O. Buzbas, D. A. Petrov, Recent Selective Sweeps in North American Drosophila melanogaster Show Signatures of Soft Sweeps. *PLoS Genet.* **11**, e1005004 (2015).

103. K. E. Lotterhos, *et al.*, Composite measures of selection can improve the signal-to-noise ratio in genome scans. *Methods Ecol. Evol.* **8**, 717–727 (2017).

104. S. Purcell, *et al.*, PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

105. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).

106. S. Anders, P. T. Pyl, W. Huber, HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

107. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

108. S. S. Mangiafico, An R companion for the handbook of biological statistics. *Available: rcompanion. org/documents/RCompanionBioStatistics. pdf. (January 2016)* (2015).

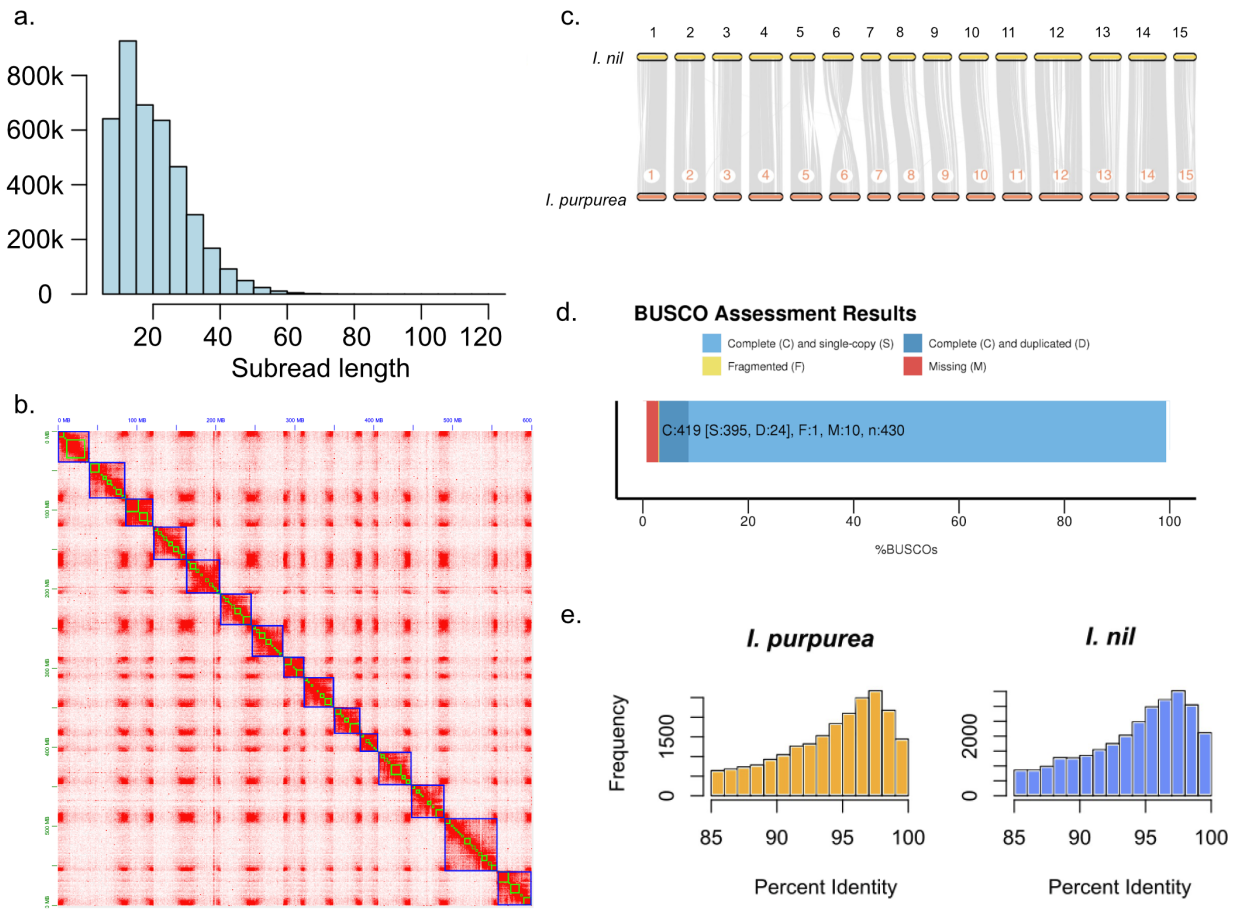
109. R. C. Team, R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2015).

110. J. Fox, S. Weisberg, *An R Companion to Applied Regression* (SAGE Publications, 2018).

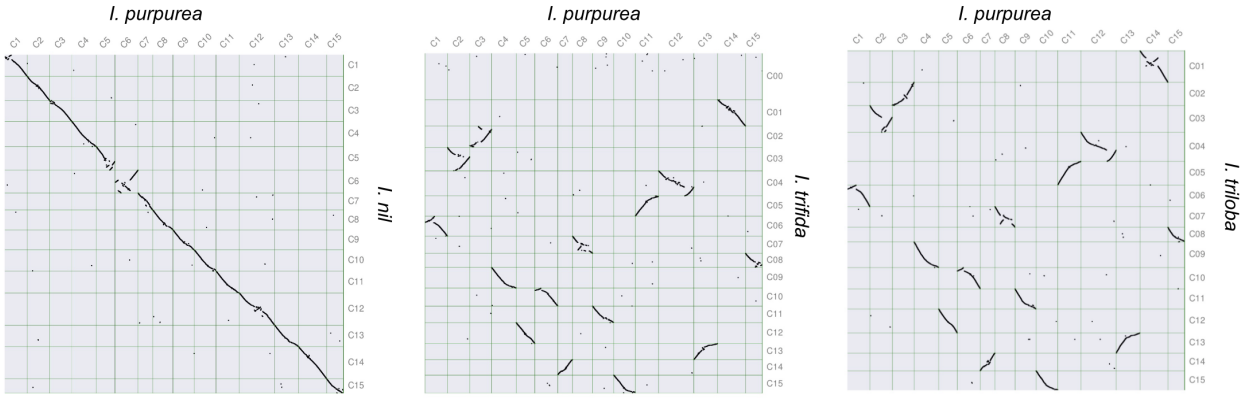
111. R. V. Lenth, Least-Squares Means: The R Package lsmeans. *Journal of*

Statistical Software, Articles **69**, 1–33 (2016).

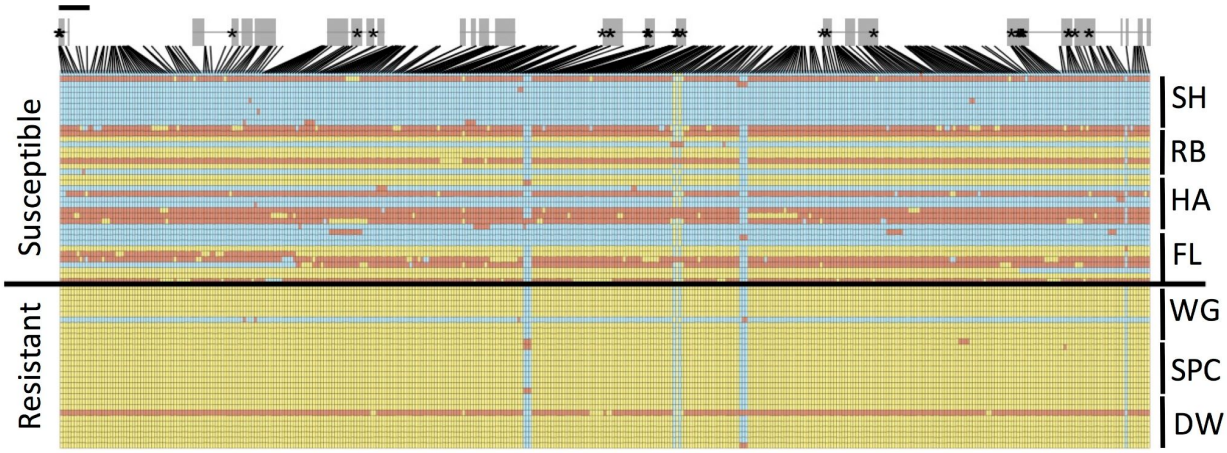
Supplemental Figures



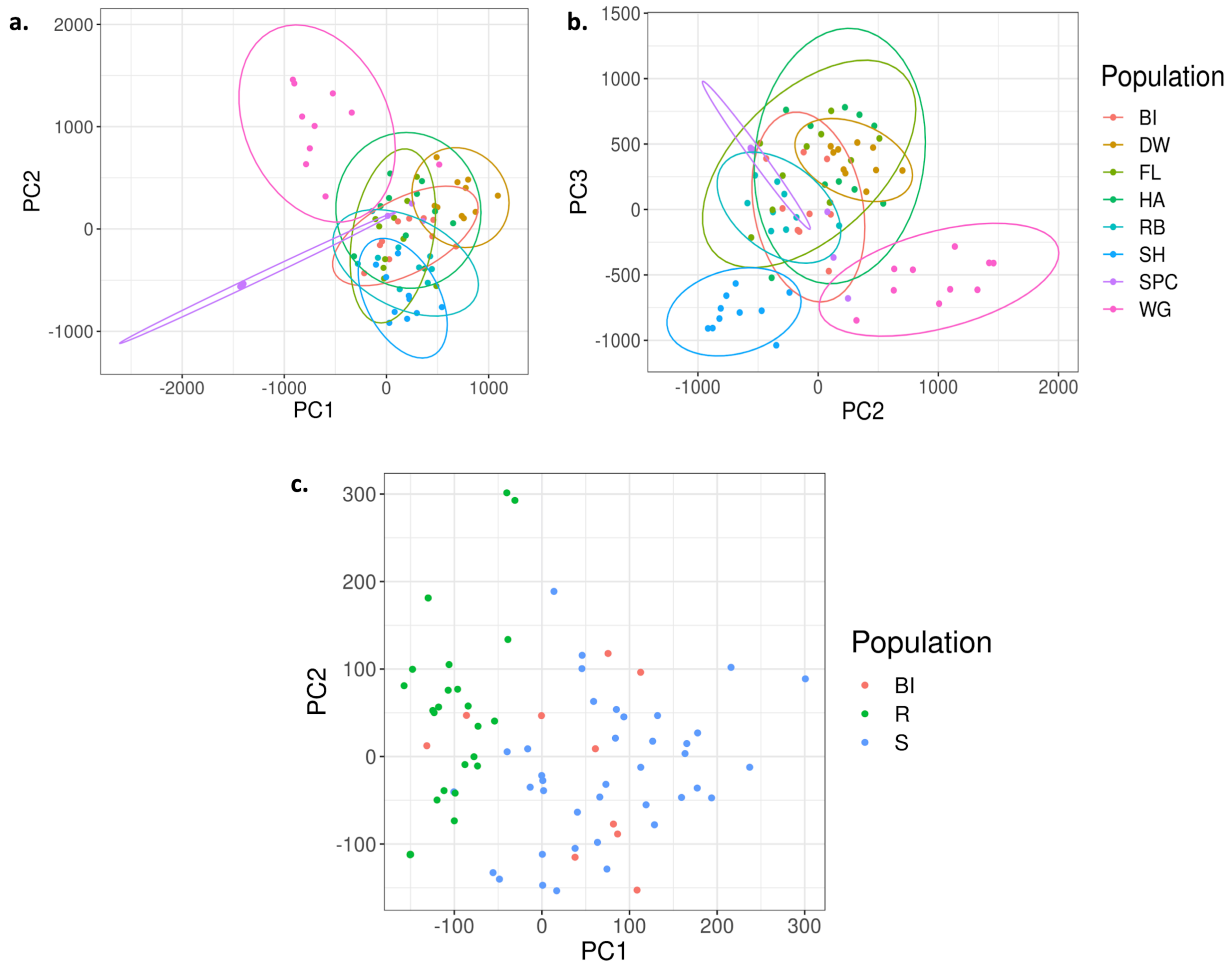
Supplementary Figure S1: Chromosome scaffolding and renaming. a) Raw PacBio Sequel filtered subread lengths. b) Phase Genomics Hi-C Proximo scaffolding results produces 15 chromosome pseudomolecules. c) Synteny of this *Ipomoea purpurea* assembly against *Ipomoea nil* was used to orient and name *I. purpurea* pseudomolecules. d) BUSCO results against Viridiplantae odb10 indicate high completeness of conserved gene sets in the raw assembly. e) Retrotransposon annotations using LTRharvest were used to compute pairwise LTR identities within LTRharvest for both *I. purpurea* and *I. nil*.



Supplementary Figure S2: Synteny of the *I. purpurea* genome against related Convolvulaceae species, including *I. nil*, *I. trifida*, and *I. triloba*.



Supplementary Figure S3: Signs of selection across conserved haplotype of multiple glycosyltransferases for each individual on Chromosome10. Exons are shown in grey. Blue and yellow indicate homozygotes, red indicates heterozygotes; stars indicate non-synonymous substitutions. Black bar above gene models indicate 1kb.



Supplementary Figure S4: Individuals from the sampled populations do not cluster into distinct resistant and susceptible groups when using all the SNPs (a and b), but there is some grouping when only considering the SNPs from the regions under selection (c).