bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Repeated origins, gene flow, and allelic interactions of herbicide resistance mutations in a widespread agricultural weed

Kreiner, J.M.[1*†‡], **Sandler, G.**[1], **Stern, A.J.**[2], **Tranel, P.J.**[3], **Weigel, D**[4], **Stinchcombe, J.R.**[1§], **Wright, S.I.**[1§]

***For correspondence:** julia.kreiner@mail.utoronto.ca (JMK)

[§]These authors contributed equally to this work

Present address: [†]Department of Botany, University of British Columbia, Vancouver, Canada; [‡]Biodiversity Research Centre, University of British Columbia, Vancouver, Canada

[1]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada; [2]Graduate Group in Computational Biology, UC Berkeley, Berkeley, USA; [3]Department of Crop Sciences, University of Illinois at Urbana–Champaign, Urbana, USA; [4]Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tübingen, Germany

Abstract Causal mutations and their frequency in nature are well-characterized for herbicide resistance. However, we still lack understanding of the extent of parallelism in the mutational origin of target-site resistance (TSR), the role of standing variation and gene flow in the spread of TSR variants, and allelic interactions that mediate their selective advantage. We addressed these questions with genomic data from 18 agricultural populations of *Amaranthus tuberculatus*, which we show to have undergone a massive expansion over the past century, with a contemporary effective population size estimate of $8x10^7$. We found nine TSR variants, three of which were common—showing extreme parallelism in mutational origin and an important role of gene flow in their geographic spread. The number of repeated origins varied across TSR loci and generally showed stronger signals of selection on *de novo* mutations, but with considerable evidence for selection on standing variation. Allele ages at TSR loci varied from ~10-250 years old, greatly pre-dating the advent of herbicides. The evolutionary history of TSR has also been shaped by both intra- and inter-locus allelic interactions. We found evidence of haplotype competition between two TSR mutations, their successes in part modulated by either adaptive introgression of, or epistasis with, genome-wide resistance alleles. Together, this work reveals a remarkable example of spatial parallel evolution—the ability of independent mutations to spread due to selection contingent on not only the time, place, and background on which they arise but the haplotypes they encounter.

Introduction

The evolution of resistance in agricultural pest populations occurs rapidly and repeatedly in response to herbicide and pesticide applications. Reports of herbicide resistance across agricultural landscapes have been steadily growing, threatening crop productivity and greatly raising costs for agricultural production (***Peterson et al., 2018***). These reports put a lower limit on the estimated number of unique resistance cases—of over 500 across the globe—based on just the occurrence of resistance to different herbicide mode-of-actions across different species (***Heap, 2014***) and barring the probably minor role of interspecific hybridization. For acetolactate synthase (ALS) inhibiting herbicides alone, over 160 species have evolved resistance since the first report of resistance in 1986, which was only five years after their initial introduction (***Comai and Stalker, 1986***; ***Whitcomb, 1999***; ***Heap, 2014***). These numbers are likely a vast underestimate of the repeatability of herbicide

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

resistance evolution. For ALS herbicides, for example, non-synonymous substitutions at 8 distinct codons confer resistance (**Tranel and Wright, 2002**), with multiple independent causal mutations often occurring in the same population (**Heap, 2014**; **Kreiner et al., 2018**). In addition to repeated resistance evolution through distinct causal resistance loci, it is likely that for a single locus, resistance mutations have arisen repeatedly within a species (**Kreiner et al., 2019**). While these observations suggest herbicide resistance may be among the most extreme cases of contemporary parallel evolution, it still remains unclear how often resistance is spread across the range through gene flow versus repeated independent origins.

Population genomic approaches can greatly help to understand the origin and spread of herbicide resistance. Genomic methods have tested for differences in population structure among resistant and susceptible agricultural populations (**Küpper et al., 2018**), reconstructed complex genomic regions associated with resistance (**Molin et al., 2017**), and investigated patterns of selection on and the extent of convergence between loci conferring non-target site resistance (**Van Etten et al., 2019**; **Kreiner et al., 2020**). But even for validated resistance mutations that occur within the gene whose product is targeted by the herbicide (target-site resistance, TSR mutations), investigations of their recent evolutionary history are sparse (but see **Flood et al.** (**2016**); **Kreiner et al.** (**2019**)). With large-effect mutations identified as being causal for conferring target-site resistance to nine herbicides at 19 loci across many species (**Murphy and Tranel, 2019**), the field is ripe for the application of population genomic techniques for resolving the evolutionary history of herbicide resistance in weed populations.

In contrast to most of the selective sweep literature coming from within-host studies of drug resistance in HIV (e.g. **Pennings et al.** (**2014**); **Feder et al.** (**2016**))—where sweeps occur in a closed-system, often starting from a single founding viral lineage and evolving within individual patients—evolutionary patterns of resistance to herbicides across a relevant agricultural landscape are by no means expected to be as tidy (but see **Feder et al.** (**2017**, 2019) for spatial structure in HIV evolution). Weedy agricultural populations themselves, or at least genotype compositions, may be transient in space and time due to widespread gene flow by pollen and seed dispersal, and changing selection regimes through rotations of both focal crops and herbicide applications and mixtures (**Naylor, 2003**; **Holst et al., 2007**; **Neve et al., 2009**). Consequently, persistent agricultural weed populations likely comprise a collection of resistant haplotypes that have arisen and dispersed across the landscape. Recently, population genomic evidence supported this prediction for a subset of newly problematic glyphosate-resistant agricultural populations of *Amaranthus tuberculatus* in Ontario, Canada, where both genome-wide and fine-scale genomic signatures in and around an EPSPS gene amplification (conferring resistance to glyphosate herbicides) shows similarity to many distinct populations in the Midwestern US (**Kreiner et al., 2019**), suggesting long-distance dispersal as contributing to the spread of glyphosate resistance. In a metapopulation context, problematic weeds are likely to have large effective population sizes (N_e) as a result of large census population sizes, widespread distributions, considerable seedbanks (**Hedrick, 1995**; **Vitalis et al., 2004**), and occasionally outcrossing mating systems, offering more independent backgrounds on which new mutations arise (**Kreiner et al., 2018**). Thus, large, outcrossing weed populations such as agricultural *Amaranthus tuberculatus* populations may be characterized by both multiple origins of TSR mutations and haplotype sharing among a network of populations across the landscape.

The evolution of resistance to a given herbicide in a network of interconnected populations should be well described by a model of spatial parallel mutation (**Ralph and Coop, 2010**), where the spread of new adaptive mutations across the range depends on the mutation rate, the extent of gene flow among subpopulations, and the density of adaptive haplotypes already present in those subpopulations. How independent mutations interact upon meeting will depend on the extent of their intra- and inter-locus interactions (**Ralph and Coop, 2010**), and in the case of resistance, their relative cost and benefits in the absence and presence of herbicides (**Vila-Aiub et al., 2009**; **Baucom, 2019**).

Here we investigate repeatability in the evolutionary histories of characterized target-site resis-

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

tance mutations for acetolactate synthase (ALS) inhibiting and protoporphyrinogen oxidase (PPO) inhibiting herbicides in a widespread agricultural weed, *Amaranthus tuberculatus*. We infer the number of TSR mutational origins across populations, examining the signals left behind by both mutation and recombination. Specifically, we implement ancestral recombination graph (ARG) based methods that offer a powerful approach for inference of selective history by providing near-complete information on relatedness among haplotypes (***Rasmussen et al., 2014***), and coupled with tree-sequenced based estimates of (N_e) through time, allow for powerful hypothesis testing about the role of standing variation versus new mutation in local adaptation. ARGs have seen limited implementation outside of human systems for examining patterns of local adaptation but recently have been used to infer the evolutionary processes that govern islands of differentiation across birds (***Hejase et al., 2020***).

In addition to inferring the number of mutational origins, we characterize evidence of selection on new mutations versus standing genetic variation on each origin directly by quantifying heterogeneity in their allelic ages based on contemporary N_e estimates, implementing two tree-based methods that tests for a consistent or a recent shift in selection throughout a mutation's history, and inferring selection coefficients through time. We find considerable heterogeneity in whether independently arisen TSR variants are associated with pronounced signals of selection, some of which may be mediated by intra- and inter-locus allelic interactions. We thus also examine these interactions directly, in order to assess the extent that extreme selection from herbicides on TSR mutations has led to observable evidence of genome-wide adaptive introgression or epistasis, as well as competition among resistant haplotypes. Our detailed population genomic analysis of the repeatability of target-site herbicide resistance evolution advances our understanding of rapid adaptation of multicellular organisms to an extreme selective pressure.

Results

Types of Target-site Mutations

We examined eight ALS, three PPO, and one photosystem II protein D1 (psbA) loci for substitutions known to confer TSR to 3 distinct classes of herbicides in *Amaranthus tuberculatus* and used whole-genome information to determine signatures of selection and demographic patterns in 19 agricultural populations in the Midwestern USA and Southwestern Ontario, Canada. Across 152 individuals, we found nine types of known TSR mutations to two classes of herbicides for which the causal alleles have been verified (Table 1). These nine mutations include one mutation conferring resistance to PPO inhibiting herbicides, and eight distinct ALS resistance mutations. We previously characterized two types of glyphosate resistance (SNP and gene amplification) in these samples (***Kreiner et al., 2019***), so we focus here primarily on PPO and ALS TSR mutations.

The nine unique PPO and ALS target-site resistant mutations occur at seven distinct amino acid positions, with two positions segregating for multiallelic resistance: two non-synonymous changes at codons 197 and 653 in the ALS gene. Six out of nine variants are rare (<5 instances found), in contrast to the common Trp-574-Leu and Ser-653-Asn nonsynonymous substitutions, and one variant of the ΔGly210 deletion (Table 1). Notably, the most common resistance mutational variant (referring to identity-by-state), Trp-574-Leu, is found in 53% of agricultural individuals, the second most common, Ser-653-Asn, in 32% of individuals (Table 1). Between these two most frequent ALS mutations alone, 74% of individuals sampled here segregate for resistance to ALS-inhibiting herbicides. Accounting for rare ALS resistance mutations only increases this percentage to 75%, since these rare SNPs are almost exclusively found stacked on other resistant backgrounds.

At a regional scale (within Essex County, Walpole Island, and the Midwestern USA), multiple causal changes segregate within each region for ALS resistance. At the population level, 5/8 populations within the Midwest, 5/5 populations in Essex County, and 4/6 populations in Walpole harbour multiple causal ALS mutations (Table 1). Thus, just at the level of resistance variant types, we observe convergent evolution of resistance to ALS herbicides at national, regional, and population

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Table 1. Number (and frequency) of resistant individuals and alleles for loci causal to PPO and ALS herbicides globally, and within each agricultural region.

	PPO	**ALS**						
	ΔGly210	Trp-574-Leu	Ser-653-Asn	Ser-653-Thr	Gly-654-Phe	Pro-197-Leu	Pro-197-His	Ala-122-Ser	Asp-376-Glu
Global Indiv.	22 (0.145)	80 (0.526)	48 (0.316)	2 (0.013)	1 (0.007)	1 (0.007)	1 (0.007)	2 (0.013)	3 (0.020)
Global Allele	25 (0.082)	106 (0.349)	59 (0.194)	2 (0.007)	1(0.003)	1 (0.003)	1 (0.003)	2 (0.007)	3 (0.010)
Walpole Allele	0	12 (0.162)	12 (0.162)	1 (0.013)	0	0	0	0	0
Essex Allele	1 (0.013)	23 (0.288)	35 (0.438)	0	1 (0.013)	0	1(0.013)	2 (0.025)	3 (0.038)
Midwest Allele	24 (0.16)	71 (0.47)	12 (0.08)	0	0	1 (0.007)	0	0	0

scales.

Regional Selective Sweep Signals

To learn how and how often the individual variants might have arisen, we first visualized selective sweep patterns at PPO and ALS genes—two genes that happen to occur 250 bps apart—and especially with respect to the common Trp-574-Leu, Ser-653-Asn, ΔGly210 loci. In particular, we assayed the extent to which strong, recent selection from herbicides at these genes has led to reductions in diversity, and increases in haplotype homozygosity and linkage in the neighbouring genomic region. Corresponding selective sweep signals appear to be highly heterogeneous across geographic regions and across TSR mutations (Figure 1). The most pronounced selective sweep signal at the regional level is for the ALS Ser-653-Asn locus, in our large collection of nearby populations from Essex county. These resistant haplotypes show a dramatic excess of homozygosity over susceptible haplotypes for nearly 10 Mb (XPEHH, Figure 1 top-left). The breadth of the impact of selection on local chromosome-wide LD is worth noting—this extended sweep signal is even larger than what was seen for an EPSPS-related gene amplification whose selective sweep in response to glyphosate herbicides spanned 6.5 Mb in Walpole populations (***Kreiner et al., 2019***). In comparison to homozygosity in Essex, Ser-653-Asn haplotypes in Essex show a more muted signal in LD with missense SNPs (r^2) and pairwise diversity for a shorter ~6 Mb (Figure 1, top middle/right). As a quick aside, we used missense SNPs as we expected them to be less influenced by population structure and admixture (***Good, 2020***) compared to synonymous SNPs, but present the correlation between genome-wide LD with synonymous and missense SNPs in Sup Figure 3.

Outside of Essex, and when excluding Ser-653-Asn, selective sweep signals are much subtler in Walpole and especially in Midwest, the latter in part likely driven by lower regional allele frequencies and our broader population sampling (Table 1). For the other common ALS TSR allele, LD between Trp-574-Leu and surrounding missense mutations remains elevated for ~6 Mb over the chromosome-wide background in both Essex and Walpole (r=0.0630 and r=0.0461 respectively, compared to 0.0300 in Midwest) (Figure 1), but haplotypes associated with Trp-574-Leu show little reduction in diversity and homozygosity, except for the latter in Walpole. The PPO ΔGly210 is only found at considerable frequencies in the Midwest, but regional sweep signals based on homozygosity, diversity, and LD are absent with respect to presence absence for the deletion.

Despite inconsistent sweep signals, the mutations we describe here are extremely likely to have experienced selection over their history, but varying over space and time. We know from previous functional validation that these mutations are causal for resistance to ALS or PPO inhibiting herbicides [in *Amaranthus tuberculatus* for the PPO deletion, as well as ALS Trp-574-Leu, and both Ser-653-Asn and Ser-653-Thr substitutions (***Matthew J. Foes et al., 1998***; ***Shoup et al., 2003***; ***Patzoldt and Tranel, 2007***), and in other congeners for the remaining mutations (***McNaughton et al., 2001***; ***Whaley et al., 2004***; ***Nakka et al., 2017***; ***Singh et al., 2018***)]. ALS inhibiting herbicides

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

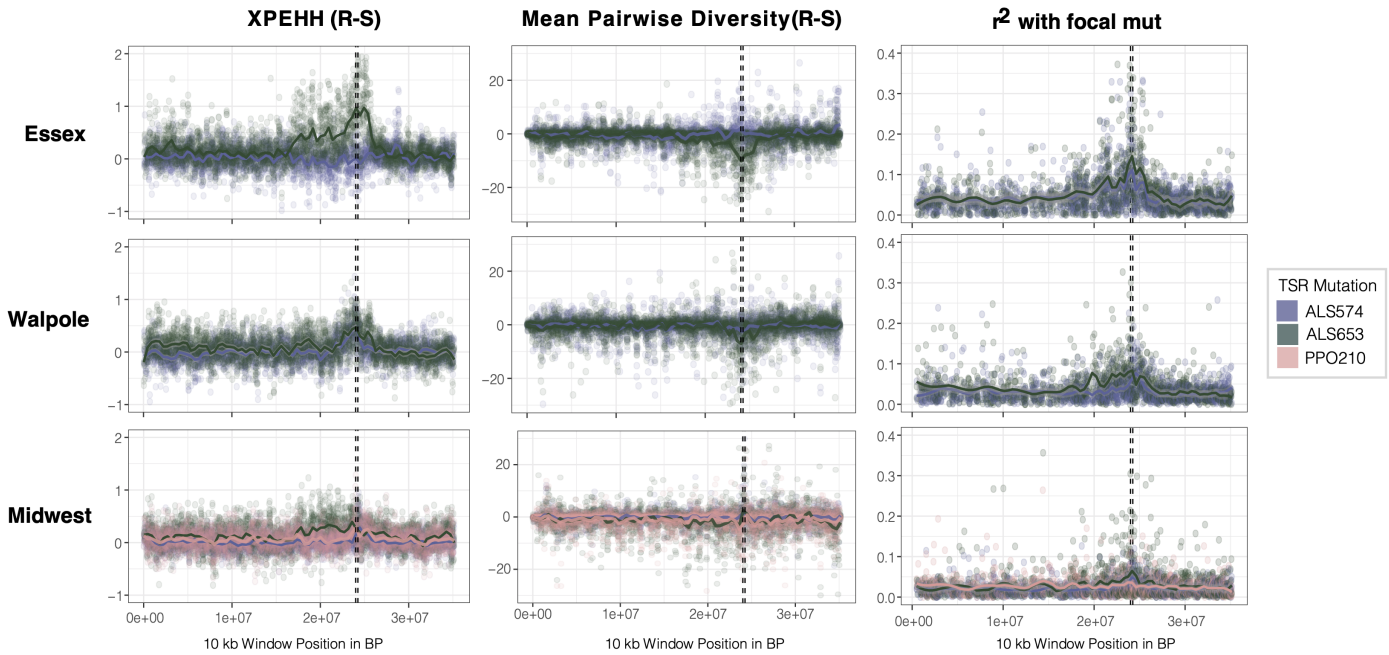


Figure 1. Sweep-scan summary statistics by geographic region. Left) Difference in integrated haplotype homozygosities (XPEHH) between haplotypes carrying the focal TSR mutation and susceptible haplotypes. Middle) Difference in mean pairwise diversity between haplotypes carrying the focal TSR mutation and susceptible haplotypes. Right) r^2 of other missense mutations with focal TSR mutation. In all columns, dashed vertical lines denote PPO (left) and ALS (right) genes, their close proximity of 250 kb apparent here.

were among the most popular mode-of-action for weed control in crops beginning in the 1980s (***Brown, 1990***) and are widely used in both corn and soy production systems, although their prevalence decreased in the 1990s as resistance became common and with the widespread adoption of glyphosate-resistant crops. When at its peak in the late 1990s, PPO herbicides represented 10% of annual applications in the USA; by 2006 its national application rate had dropped to 1%. However, following widespread resistance to glyphosate, there was a resurgence in use of PPO inhibitors both for soil-residual and postemergence control of *A. tuberculatus*, particularly in soybean (***USDA-NASS, 2012***; ***Dayan et al., 2018***). To better understand the evolutionary processes that govern the spread of resistance within and among populations, we next characterized the number of independent origins for each high-frequency TSR variant and attempted to identify genomic causes and consequences of their distinct selective histories.

Inferring the genealogical history of TSR mutations

The patterns of similarity among phased haplotypes (including 1 kb upstream and downstream of both the ALS and PPO genes) across resistant individuals illustrates that repeated origins may be the rule rather than the exception for common resistance alleles (Sup Figure 1; ALS Trp-574-Leu and Ser-653-Asn, PPO ΔGly210); all three common resistance types show multiple clusters of resistant haplotypes spread across their two respective gene trees. Nevertheless, very few origins (defined by identity-by-descent) are specific to a given geographic region, and fewer still population-specific, especially for origins of ALS Trp-574-Leu. The absence of geography-specific origins in resistance types highlights the potential importance of gene flow in the spread of resistance mutations across populations. A gene tree based on raw pairwise differences between haplotypes, as illustrated here, sets an upper limit on the number of independent origins for each mutation—from this plot alone, one could infer four origins of the ALS Ser-653-Asn type, at least eight of the Trp-574-Leu type, and three of the PPO ΔGly210 type. However, recombination events will break apart the association between a new resistance mutation and its original haplotypic background, potentially

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

inflating the number of independent resistance clusters mapping across the tree.

Because recombination causes ancestral haplotypes to decay in size as they are passed down through time, linked sites may not necessarily have identical genealogies. Nearby sites may vary in the time to their most recent common ancestor, and thus inferring the full genealogical history of two or more genomes requires inference of the coalescent history of recombinational units across the genome. Ancestral recombination graphs (ARGs) generalize the inference of coalescent history along a recombining unit, allowing for more accurate inferences of demographic and selective history (***Hudson, 1983***; ***Griffiths and Marjoram, 1996***, ***1997***).

We inferred the ARG for 20,000 SNPs encompassing both ALS and PPO genes (a ~1 and ~10 kb gene respectively, 250 kb apart from one another) using the program ARGweaver (***Rasmussen et al., 2014***; ***Hubisz and Siepel, 2020***). We then used arg-summarize to extract the tree corresponding to each focal TSR locus that maximizes the likelihood of our data given the model across 500 MCMC samples after burn-in (Sup Figure 2). For all three common TSR variants, ALS Trp-574-Leu, ALS Ser-653-Asn, and the PPO ΔGly210 deletion, we found evidence for multiple independent mutation events producing the same variant. In every one of the 500 MCMC samples of the ARG, trees at each of the three focal TSR loci were consistent with a finite sites model, driven by multiple mutational origins being the most parsimonious explanation for haplotype clustering across the tree and across iterations of the ARG algorithm (Figure 2A). Our data suggest that the most common resistance mutation, ALS Trp-574-Leu, arose six times independently, PPO ΔGly210 three times, and ALS Ser-653-Asn twice (Figure 2A; ignoring singletons and polytomies at the base of the tree). It is worth noting that, as expected, these numbers are smaller than what was inferred from the pairwise distance gene tree approach, especially for the ALS Trp-574-Leu and ALS Ser-653-Asn variants, suggesting that some of the ALS TSR mutations are old enough that recombination has played a role in further segmenting their founding haplotypes.

When haplotypes belonging to distinct mutational origins are mapped across populations (Figure 2B), it is clear that, despite the many independent mutational origins, gene flow has also played a major role in the spread of resistance across the landscape. Only two mutational origins are population-specific—Trp-574-Leu #2 and PPO ΔGly210 #11—both representing the lowest frequency origins for a given TSR locus. In contrast, haplotypes from the three most common origins of TSR to ALS herbicides—Trp-574-Leu #1 and #3, and Ser-653-Asn #7 (corresponding to 39, 25, and 47 haplotypes, respectively)—can be mapped to 10, 10, and 12 populations. While each of these three common mutational origins can be found in both the Midwest US and Ontario, there is clear stratification in terms of how common these alleles are across these two regions, as would be expected from an initially localized origin. Both ALS574 mutations are found at much higher frequencies in Midwestern populations (89% and 64% of haplotypes being found in the Midwest), whereas the ALS653 #7 mutation is near exclusive to Ontario populations (96% found in either Essex or Walpole), except for two resistant haplotypes found in Missouri. Only one mutational origin is fully unique to Ontario populations, ALS574 #5, whereas ALS574 #2 and #6, as well as PPO ΔGly210 #9 and #11 are unique to the Midwest. With the expectation that the geographic distribution and frequency of these 11 distinct mutational origins should in part reflect both their age and the strength of selection they have experienced over their history, we next used tree-based approaches for exploring the extent of heterogeneity in selective signals and allelic ages.

We initially performed a tree-based test of non-neutral allele frequency change over two distinct timescales to approximate scenarios of selection on standing variation versus *de novo* mutation. These tests focus on characterizing signals of selection across either a very recent proportion of the tree or over the entire time since the mutation first arose. This approach avoids explicit estimates of the absolute age of mutations, since those depend heavily on accurate N_e estimates through time. We approached these tests of selection one unique mutational origin at a time, excluding all other resistant lineages from the tree, such that our estimates of the probability of selection for a given mutational origin is relative to coalescent events of susceptible lineages. Selection should lead to high-fitness lineages disproportionally giving rise to offspring in the next generation. As

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

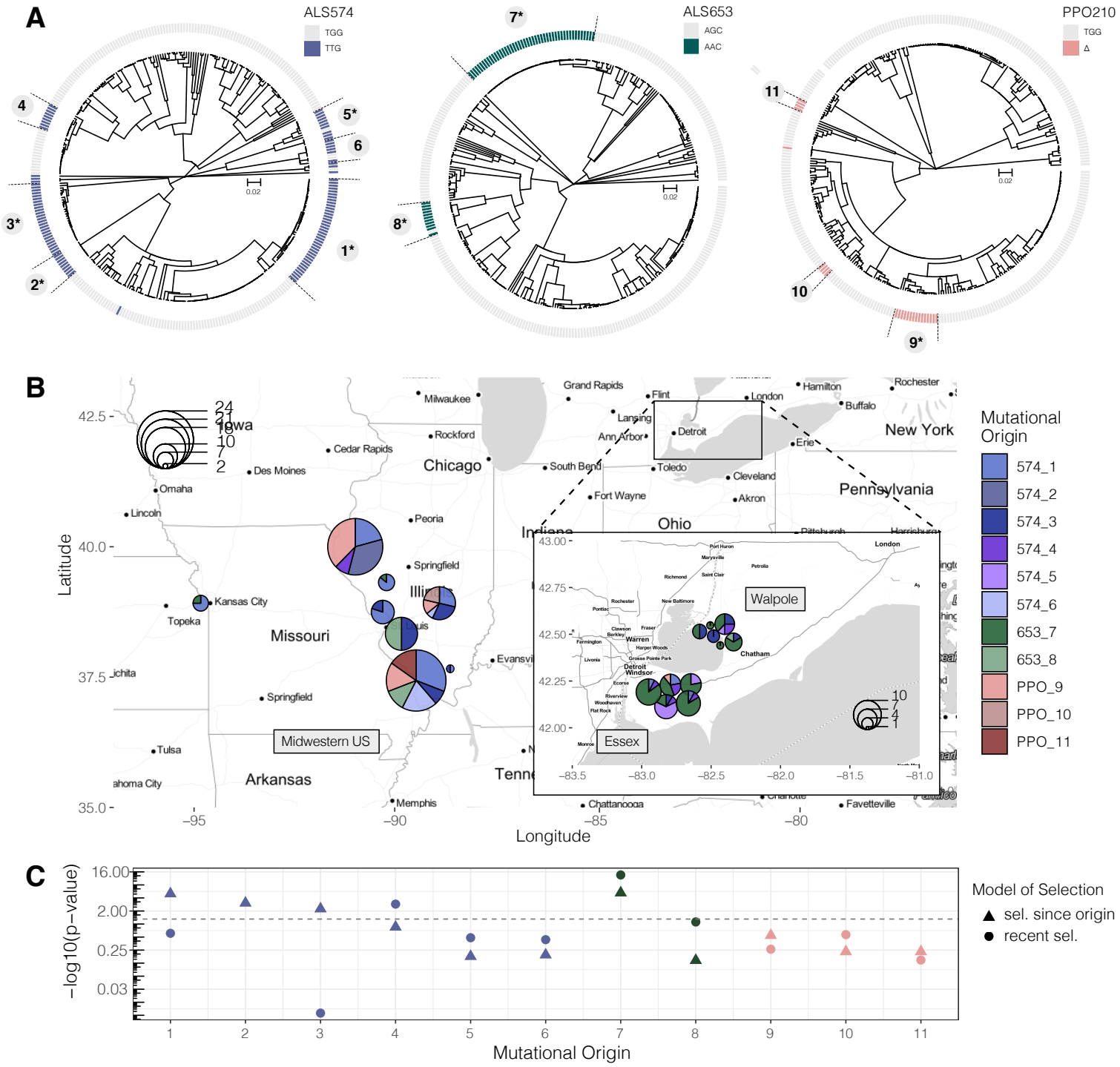


Figure 2. Repeated independent origins of three target-site resistance variants across the range and likelihood of selection under new mutation versus standing variation. **A**) Trees corresponding to ARGs estimated at focal TSR loci. Numbers around trees identify clusters of resistant haplotypes consistent with an independent origin. The presence of an asterisk at each origin number implies significant evidence of selection since the mutation arose *de novo* at $p < 0.05$ against the null distribution. **B**) Map of mutational origins inferred from trees extracted from ARGs at the focal TSR loci. Mutational origins are by and large shared across populations, although regions show clear differences in the frequency of each mutation. **C**) p-values (-log10) of tree-based tests of selection on each mutational origin of TSR under two alternative models of selection; selection on a mutation since its origin (approximating selection occurring *de novo* mutations) versus selection occurring only recently (approximating selection on standing genetic variation). The horizontal dashed line denotes the *p*-value cutoff of α=0.05.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

such, to test for the scenario of consistent selection over the entire course of an independent mutational event's history, we evaluated whether there was evidence of a particularly extreme increase in the frequency of that mutation since it first arose, beyond the neutral expectation (testing the scenario of selection on a *de novo* mutation). To test the scenario of selection on an independent mutation that had been standing before the onset of selection, we similarly evaluated evidence of an extreme increase in the frequency of a given mutational origin, but over a timescale that samples just its most recent history (0.02% of our tree). Those mutations that showed evidence of a selection over the more recent timescale but not since their origins represent candidates for selection on standing variation. Specifically, we implemented the tree-based statistic based on order of coalescent events as proposed in (***Speidel et al., 2019***), in addition to a modified version of this statistic that evaluates the probability of selection on more recent timescales (Methods, Tree-based tests for selection).

Under the scenario of selection since the origin of the mutation, four out of 11 mutational origins we tested were significant after a 5% false discovery rate (FDR) correction (Figure 2C). Since it arose, the Ontario high frequency ALS653 #7 variant showed the strongest signature of selection across all origins and all TSR loci (*p*-value= 5.58e-06), followed by three Midwestern high frequency mutations: ALS574 #1 variant (*p*-value= 9.81e-06; 90% found in the Midwest), ALS574 #2 variant (*p*-value= 8.85e-04; 100% found in the Midwest) and ALS574 #3 variant (*p*-value=0.00565; 64% found in the Midwest). In contrast to other origins at the same loci showing some of the most extreme evidence for selection since their origin, ALS653 #8 and ALS574 #4, #5, and #6 variants lacked such a signal. Furthermore, no PPO mutation showed evidence of consistent selection since they arose.

In contrast, a test for selection over the most recent 0.02% of the tree showed that only two out of 10 variants that originated before this period had significant evidence of selection over this recent timescale, after FDR correction (Figure 2C). Both of these showed much stronger evidence for recent selection, one of which (ALS653 #7) also had significant evidence for selection occurring since the mutation arose—implying that while it is a strong candidate for selection on a *de novo* mutation that it has also experienced particularly strong recent selection. We were unable to test one variant, ALS563 #2, for this scenario as it predated our recent timescale threshold, but with significant evidence of selection since it arose it clearly displays signals of particularly recent selection.

N_e through time, age of TSR alleles, and selection coefficients

A more direct assessment of the role of resistance adaptation from standing genetic variation or new mutation comes from estimates of allele age relative to the onset of the selection pressure, which depends greatly on the accuracy of N_e estimates over the relevant evolutionary timescale. Namely, for herbicide resistance evolution, we posit that the relevant N_e is most likely the N_e over the last half-century or less, corresponding with the introduction of agronomic pesticide regimes. While we had previously used δaδi (***Gutenkunst et al., 2009***) to infer species-wide effective population size (where N_e~500,000; ***Kreiner et al.*** (***2019***)), here we used Relate to infer effective population size through time from genome-wide tree sequence data (***Speidel et al., 2019***, ***2021***). Historical N_e between 100 and 1,000,000 years ago appears to have stayed relatively consistent, with a harmonic mean of 63,000 (SE = 7,000). Our samples show evidence for massive recent population expansion over the last 100 years, with the contemporary geometric mean N_e estimate 3 to 4 orders of magnitude higher than the historical N_e (Figure 3A). Over the timescale of ALS herbicide use (40 years), the geometric mean N_e estimate is 80,000,000.

Based on our contemporary N_e estimates relevant to the timescale of herbicide use, we rescaled allelic ages for distinct mutational origins across our ARG-inferred trees, accounting for variation across MCMC ARG samples. On average across the last 200 converged MCMC samples (Sup Figure 2) and according to haplotype groupings from the most likely ARG, we found remarkable differences in the estimated age of resistance mutations with respect to both distinct origins of the same mutation and across the three different TSR loci (Figure 3B). Estimates of the origin of the

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

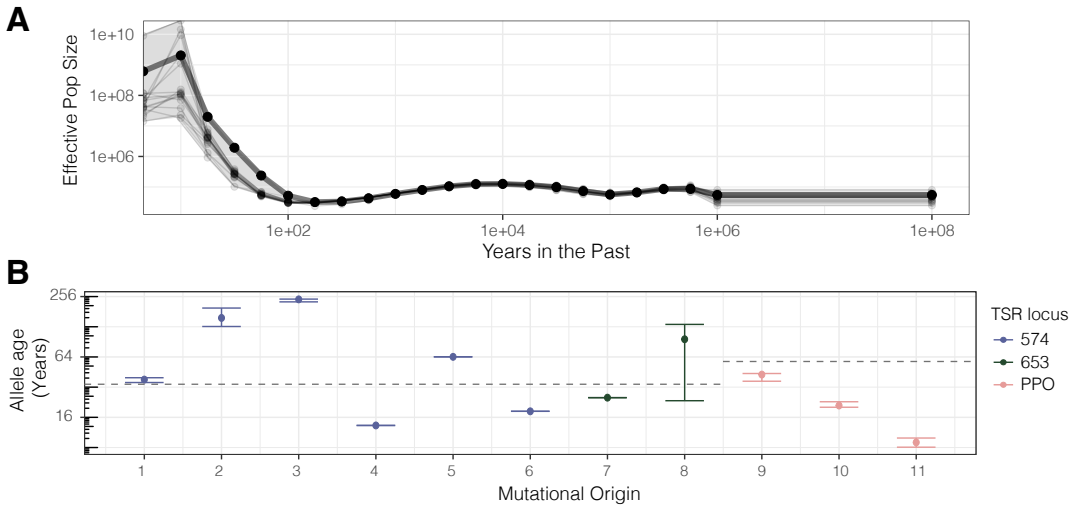


Figure 3. Contemporary population expansion and corresponding ages of TSR variants. **A**) Relate-inferred effective population size through time, illustrating a remarkable population expansion occurring over the last 100 years. The bold line indicates results from genome-wide SNPs, whereas thinner lines represent results from chromosome-by-chromosome analyses, with the shaded area showing the bounds of the variance in the chromosome-by-chromosome data. **B**) Allele age inferred from the geometric mean effective population size estimate over the timescale of contemporary herbicide use (< 50 years ago, GM[Ne] = 83,294,700). Horizontal dashed lines for ALS574 and ALS653, and PPO ΔGly210 represent the approximate onset of ALS and PPO herbicide use, respectively.

PPO ΔGly210 deletion are the most recent and least variable compared to ALS574 and ALS653 variants (mean PPO ΔGly210 age = 24 years [SE=9.9]; ALS653 = 61 [SE=35]; ALS574 = 88 [SE=37]). The younger mutational ages of PPO ΔGly210 variants and earlier but more stagnated use of PPO herbicides may imply that these mutations have experience longer periods of negative selection and/or that these resistance-conferring deletions are more deleterious compared to the more common ALS574 and ALS653 variants—consistent with the lack of net positive selection detected for PPO ΔGly210 mutations.

For the origin with the strongest evidence of selection, we additionally estimated the associated selection coefficient through an approximate full-likelihood test based on coalescent rates from the tree sequence, which explicitly takes into account changing effective population sizes through time (***Stern et al., 2019***). Over the last 30 years, the ALS653 #7 origin is consistent with an s = 0.026 (LR = 6.80, χ^2 = 13.56, p-value=2.3e-4) within Ontario (corresponding to an allele frequency = 0.29). If we constrain this test to just the last 10 years, this estimate increases substantially to s = 0.072, consistent with evidence from the tree-based tests over two timescales for particularly recent selection on this origin. The magnitude of this fitness advantage conferred from a single locus implicates herbicides as a powerful force for shaping genome-wide diversity throughout the species range, especially when considering the role of not only temporally, but also spatially varying selection along with rotating crop and herbicide use, and ecological and environmental variation (e.g. imprecise herbicide applications, variation in germination time) that may prevent a portion of individuals in a population from being exposed to applied herbicides.

These summaries based on mutational similarity and shared evolutionary history have uncovered evidence of an extremely strong selective event over timescales more recent than the origin of the ALS653 mutation in Ontario. We have also characterized the spread of numerous independent origins of ALS574 mutations across agricultural regions, including variants that have been around for nearly 250 years to ones that have arisen only in the last decade. However, once spreading adaptive alleles come into contact, their path to fixation depends on the extent of their epistatic interactions (***Ralph and Coop, 2010***), selective interference, and direct benefits and pleiotropic costs.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

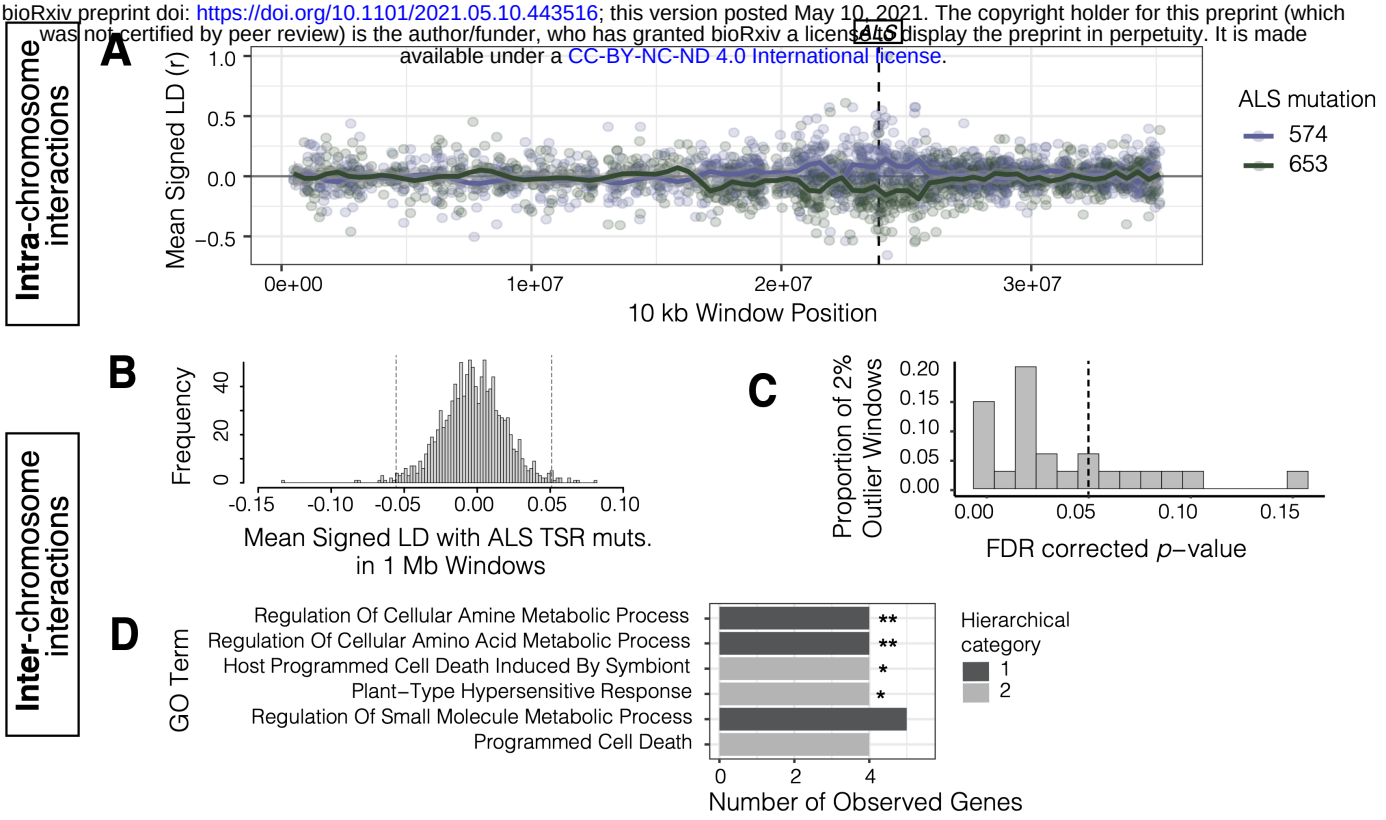


Figure 4. Signals of intra- and inter-chromosomal allelic interactions with target-site resistance mutations. **A**) Repulsion disequilibrium between two ALS target-site-resistance mutations and bi-allelic missense SNPs surrounding them on scaffold 11 in Essex. Each point shows mean signed LD in non-overlapping 10 kb windows. **B**) Distribution of mean signed LD of ALS TSR resistance mutations (ALS 574 or 653) with 1 Mb (non-scaffold 11) windows in Essex. 2nd percentile indicated by dashed vertical lines. **C**) Distribution of p-values from top 2% of genome-wide windows with most extreme LD with ALS TSR mutations, based on windowed permutations of individual-genotype assignment 1000x. **D**) Significant enrichment of GO terms for genes mapping to the 13, 1 Mb windows that show significant LD with ALS TSR mutations in Essex.

Therefore, differences in the signatures of selection based on both sweep scans and ARGs may reflect not just direct selection, but interactions among TSR mutations and with sites across the genome specific to the ancestry of local populations.

Haplotype competition and inter-locus interactions of TSR mutations

While single individuals appear to harbour both common ALS574 and ALS653 mutations, haplotype-level analyses indicate that no single haplotype harbours both mutations depiste their high frequencies (Sup Figure 1), suggesting that no recombination has occurred between these two sites and hence possible clonal interference/haplotype competition among ALS resistance haplotypes. Alternatively, recombinant genotypes creating double mutants could have strong epistatic fitness costs, leading to their selective removal. Correspondingly, in Essex, patterns of signed linkage disequilibrium (LD) of bi-allelic missense SNPs with the focal ALS574 mutation (n=19) tend to be positive, while LD of the bi-allelic missense SNPs with the focal ALS653 mutation (n=34) tend to be negative for a stretch of more than 10 Mb. Contrasting positive and negative r values for focal mutations and flanking variants, indicating linkage to alternative flanking variants, are an indication that two mutations arose on different backgrounds and that they are in repulsion. Although an excess of LD (i.e., r) is expected around a selected site for a typical hard sweep, there is no expectation for repulsion of beneficial variants with minor alleles across large stretches of a chromosome unless other evolutionary forces are acting as recombination should rapidly break down long-range associations. At its simplest, this finding is consistent with competition between herbicide resistant haplotypes in Essex (Figure 4A).

While selective interference between two TSR mutations at distinct loci appears to be an important force in the evolution of herbicide resistant weed populations, the selective advantage of a given TSR haplotype may depend on other modifier loci across the genome. In particular,

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

we might expect that individuals that have withstood many generations of herbicide applications, for example due to large-effect TSR mutations, may have also accumulated compensatory and tolerance-conferring mutations across the genome as well as mutations that reduce fitness costs in the absence of herbicide exposure (stacking of resistance alleles; ***Preston*** (***2003***); ***ffrench Constant et al.*** (***2004***); ***Petit et al.*** (***2010***); ***Busi et al.*** (***2013***); ***Kreiner et al.*** (***2020***)). Considering that haplotype competition seems to have manifested itself in patterns of signed LD (i.e. repulsion) within the ALS containing chromosome, we posited that physically unlinked modifiers of TSR resulting from either adaptive introgression or epistasis would be reflected by strong linkage with focal TSR loci.

Within Essex individuals, we calculated mean signed LD between bi-allelic missense SNPs and focal ALS574 or ALS653 mutations in 1 Mb non-overlapping windows to identify such putative interactions. We took the upper 1% and lower 1% of windows with particularly extreme signed LD with either TSR mutation, and performed a permutation test in which we randomized TSR allele assignment among individuals 1,000 times to test whether observed LD was more extreme than we might expect given the genomic structure associated with that focal window. Compared to the null expectation, the 1 Mb window with the strongest ALS TSR association showed a significant excess of positive inter-chromosomal signed LD with ALS574 (1-tailed $p < 0.0001$, $r = +0.068$) but negative of signed LD with respect to the ALS653 mutation (r = -0.132), consistent with repulsion between TSR alleles. Upon further inspection, this 1 Mb region is directly centered on a cytochrome P450 gene, CYP82D47, that has been implicated in conferring non-target site resistance in *Ipomoea purpurea* (***Leslie and Baucom, 2014***). Of the top and bottom 1% at both ends of the distribution of signed LD with these two ALS TSR mutations (24/1156 genome-wide 1 Mb windows) (Figure 4B), 13 had p-values consistent with significantly extreme LD after FDR correction with α=0.05 (Figure 4C). These 13 windows corresponded to 348 *A. tuberculatus* genes, 120 of which had *Arabidopsis thaliana* orthologs. These 120 genes were enriched for six GO biological processes belonging to two unique hierarchical categories after FDR correction, four of which were enriched even after Bonferroni correction: cellular amine and amino acid metabolic process, programmed cell death, and plant-type hypersensitive response (Figure 4D).

ALS-inhibiting herbicides disrupt biosynthesis of branched amino-acids, and a rapid response after exposure leads to amino-acid remobilization through enhanced protein degradation (autophagy) and reduced synthesis (***Trenkamp et al., 2009***; ***Orcaray et al., 2011***; ***Zulet et al., 2013***; ***Zhao et al., 2018***). That we see enrichment for multiple terms related to amino-acid metabolism and cell death (programmed cell death and plant hypersensitive response) suggests that together these genes may act to compensate for homeostatic disturbances caused by ALS exposure on the background of large-effect TSR mutations. Two particularly interesting examples from our set of genes with strong inter-chromosome LD with ALS TSR mutations are GCN2 (general control nonrepressible 2) and KIN10 (SNF1 kinase homolog 10). Both proteins have been previously identified as playing key regulatory roles in response to herbicides, with GCN2 directly involved in homeostatic tolerance to ALS and glyphosate herbicides through regulating autophagy and amino acid signalling (***Faus et al., 2015***; ***Zhao et al., 2018***). Similarly, KIN10, a key positive regulator of autophagy in *A. thaliana*, is activated in response to photosystem II herbicides (***Fujiki et al., 2001***; ***Chen et al., 2017***).

Discussion

The application of herbicides in agricultural weed management has led to one of the best-studied examples of parallel evolution in the wild, with TSR mutations to ALS-inhibiting herbicides identified in more than 150 species (***Heap, 2014***). Here, we have studied the evolution of TSR mutations at two genes, ALS and PPO, from a genome-wide perspective across a large fraction of the range of one of the most problematic weeds in the US, *A. tuberculatus*. We found rampant evidence for both independent origins and gene flow, competition among resistant haplotypes, and interaction of large-effect TSR mutations with physically unlinked alleles with resistance-linked functions.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

These results paint a picture of the rise, spread, and fate of adaptive alleles in the face of extreme selection. Below we discuss these results in the context of past studies on parallel evolution and mutation-limited adaptation, the costs and benefits of TSR mutations and their implications for selection, and the potential role of genome-wide allelic interactions in adaptation across the range.

We detected strong evidence for parallel evolution to herbicides within *A. tuberculatus* agricultural weed populations at multiple levels. Target-site mutations conferring resistance to PPO and ALS herbicides occur at seven distinct codons, and with nine distinct variants, three of which are common and themselves show evidence of having arisen repeatedly more than 10 times across our sampled populations (Table 1, Figure 2). From a mutation-limited view of adaptation, the extent of parallelism in TSR that we observe here seems particularly extreme. However, we estimate that given the North American Amaranthus estimate of $\Theta = 4Ne\mu = 0.041$ for neutral sites (***Kreiner et al., 2019***), a new TSR mutation at any of the eight adaptive ALS mutations should arise at a rate of $2N_e$—every six generations (i.e. 0.041/2 x 8 known TSR loci = 0.164 TSR mutations per generation; see also ***Charlesworth*** (***2009***); ***Karasov et al.*** (***2010***)). This rate is likely to be an underestimate given the contemporary population size should determine the mutational supply for rapid adaptation and under models of evolutionary rescue, which itself may be closer to census size than long-term estimates of N_e from neutral polymorphism (***Karasov et al., 2010***; ***Bell, 2013***; ***Neve et al., 2014***; ***Kreiner et al., 2018***). Indeed, if we modify this value to reflect the contemporary estimate of N_e over the last 50 years (~8x10^8) and assume an *A. thaliana* mutation rate of 7x10^{-9} (***Ossowski et al., 2010***; ***Weng et al., 2019***), our Θ becomes > 1 and a new mutation at any TSR codon should arise every generation—consistent with the remarkably parallel mutational origins we describe here. Thus, parallelism in herbicide resistance adaptation in *Amaranthus tuberculatus* appears to be on par with prokaryotic adaptation and pesticide resistance adaptation in *Drosophila melanogaster* where population sizes on the order of Θ ~1 facilitates adaptation to occur rapidly, without being limited by mutational input at single sites (***Karasov et al., 2010***).

In the context of such extreme recurrent evolution, we still find an important role of gene flow in the spread of herbicide resistance across the range. Not only do agricultural regions and populations within them harbour multiple origins of TSR, but distinct recombinational units harbouring these mutational origins also map to many populations (Figure 2B). In part, widespread movement of *A. tuberculatus* and TSR variants across the North American range is likely to reflect the massive recent expansion we see here (Figure 3A)—population size increasing by four orders of magnitude over the last 100 years. This expansion also corresponds well with *A. tuberculatus*'s contemporary agricultural association, suggesting that agronomic regimes are likely to have in large part facilitated the success of this weed species. Thus, both extreme mutational parallelism and a complex network of haplotype sharing, via gene flow and colonization, characterize the distribution of herbicide resistance across our sampled agricultural populations.

The role of selection on standing genetic variation in the rapid evolution of polygenic herbicide resistance has been recognized in both artificially selected and field evolved populations (***Neve and Powles, 2005***; ***Busi and Powles, 2009***; ***Busi et al., 2012***). Detrimental effects of TSR resistance mutations as a result of pleiotropic tradeoffs and fluctuating ecological selective pressures (***Lenormand et al., 2018***) have led to the question of whether such costs could be leveraged to prevent the persistence of resistance mutations (***Vila-Aiub, 2019***). We rescaled TSR allelic age estimates by the geometric mean N_e estimate over the last 50 years and found that a substantial subset of TSR mutational origins predated the onset of herbicide usage (Figure 3B). While these allele age estimates provide some of the strongest evidence for herbicide resistance adaptation from standing genetic variation to date, they only provide a first approximation given the difficulty of accounting for the monumental population expansion this species shows during tree rescaling. To account for this uncertainty, we additionally implemented a tree-based statistic that uses the order of coalescent events to infer evidence of selection since the origin of the allele (consistent with selection on a *de novo* mutation) versus selection over more recent timescales, which should be robust to population size misspecification. This test shows evidence of selection particularly on recent timescales

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

for two mutational origins (one each for ALS574 and ALS653) (Figure 2C), with one origin showing evidence only on recent timescales, consistent with selection on standing variation. Together with allele ages that predate the onset of herbicides, this implies that costs of resistance mutations are not ubiquitous, as TSR mutations can persist over longer timescales even if they are not initially favoured in the environment in which they arise.

Nonetheless, evidence for selection on *de novo* origins of resistance and allele age estimates that are younger than the onset of herbicide use tends to be more frequent than evidence for selection on standing genetic variation, suggesting that the success of TSR mutations is in large part determined by the selective regime at the time at which they arise. An exception to this is apparent in the near Ontario-specific ALS653 variant. While we estimate that this mutation has on average experienced an s=0.026 over the last 30 years, surely facilitating the persistence of this mutation, a change in environment—possibly the recent colonization of agricultural fields in Ontario, Canada—has led to its dramatic increase in frequency and relative fitness, with an $s = 0.072$ over the last decade.

Since TSR mutations can provide different ranges of protective effects in response to increasing herbicide dosage (e.g. ***Marshall et al.*** (***2010***); ***Massa et al.*** (***2011***); ***Panozzo et al.*** (***2013***)), TSR mutations should differ in their benefits under a specific crop-herbicide regime. For example, in corn and soy production systems, PPOs were typically used only in soy, whereas ALS herbicides were heavily used in both crops (***Tranel and Wright, 2002***; ***Salas et al., 2016***). Thus with typical rotations, PPO selection would occur only every other year, consistent with the absence of evidence for consistent selection on PPO TSR mutations. Furthermore, while ALS574 mutations tend to confer high levels of resistance broadly across ALS herbicides, ALS653 mutations confer resistance to fewer types of ALS-inhibiting herbicides that are also used more commonly in soy (***Patzoldt and Tranel, 2007***). This may contribute to the relatively lower global frequency of ALS653 compared to ALS574 or even suggest that the lack of rotation of both focal crops and the type of ALS herbicide led to the strong selection on the Ontario-specific ALS653 origin. Beyond locus-specific differences, our coalescent-based inference of the probability of selection on independent origins of the same variant underscores the heterogeneity in the success of a given resistance variant. Both common ALS types, Trp-574-Leu and Ser-653-Asn, show considerable variance in evidence for selection among independent origins (Figure 2C, 3B)—undoubtedly, driven by spatially varying selection, demographic events, competition with other pre-existing TSR mutations, and variable extents to which epistatic compensatory and/or tolerance conferring mutations have accumulated on their genomic backgrounds.

The outcome of parallel adaptation in a continuous species range has been thoroughly described in (***Ralph and Coop, 2010***). When the geographic spread of an adaptive mutation is migration limited, partial sweeps for parallel adaptive mutational origins that occur in distinct geographic regions will be common. However, as "waves of advance" of these distinct mutational origins expand, eventually coming into contact—as long as their selective advantage is similar, and barring epistasis between them—one beneficial allele is likely to exclude the other, with beneficial haplotypes competing along the way (***Ralph and Coop, 2010***). Given our evidence for highly parallel TSR adaptation across the range, along with widespread gene flow, we expect that this scenario fits the evolution of resistance particularly well. When beneficial alleles originate on distinct backgrounds associated with different population histories, patterns of adaptation within a population may resemble soft selective sweeps (***Ralph and Coop, 2010***). While soft-selective sweeps via gene flow appears to be a predominant pattern resulting from TSR resistance evolution, remarkably, we also find evidence of a subsequent process in such a model of parallel geographic origins, resulting from interactions among spreading TSR mutations.

Our evidence for intra-locus allelic interactions between two common TSR mutations, ALS 574 and ALS653 is in the form of a lack of recombinant double-resistant haplotypes and repulsion disequilibrium between two common resistance haplotypes (Figure 4A)—reflecting either haplotype competition, negative epistasis, or selective interference (***Hill and Robertson, 1966***; ***Gerrish***

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

and Lenski, 1998; ***de Visser and Rozen, 2005***; ***Williams and Pennings, 2020***). Populations in Essex, homogeneous for *A. tuberculatus* var. *rudis* ancestry (***Kreiner et al., 2019***), for which we have particularly fine-scale sampling, is the strongest example of this phenomenon in our data set. Likely, the mutational origin of ALS653 almost exclusive to Essex, came in on a foreign haplotype (from an unsampled population) and was swept to high frequency, competing with the selective advantage of ALS574. This interpretation is supported by the younger inferred age of ALS653 than the average age of ALS574 found in Essex, and additionally that rare (minor) missense mutations have yet to accumulate on the background of ALS653 (i.e., are in negative LD).

That we find evidence of haplotype competition (the simplest interpretation) is especially notable given the obligately outcrossing, wind-pollinated breeding system of *A. tuberculatus* (***Costea et al., 2005***). Empirical examples of a similar process, selective interference, in the literature are typically constrained to asexually reproducing organisms (in the form of clonal interference) such as *Escherichia coli* (***Hegreness et al., 2006***; ***Barrick et al., 2009***) and *Mycobacterium tuberculosis* (***Eldholm et al., 2014***; ***Sun et al., 2012***), as the lack of recombination amongst distinct beneficial lineages leads to easily observable selective interference at the level of the entire genome. Nonetheless, selective interference between resistance mutations has also been illustrated in Human Immunodeficiency Virus (HIV), which undergoes recombination at appreciable rates (***Williams and Pennings, 2020***), in the sexually reproducing *Drosophila melanogaster* (***Castellano et al., 2015***). On one hand, given the LD-based population recombination rate estimate of of $\Theta = 4N_e r$ = 0.0805 scaffold-wide and 0.0575 in a region of 100 kb on either side of ALS, it is remarkable that we recover no recombinant double ALS resistant mutant as this local Θ implies 3.7 new recombination events per generation in the distance between these two mutants (256 bp x (0.057/4)). Given the local frequencies of ALS574 and ALS653 resistance haplotype in Essex (0.29 x 0.44), 1 of these recombination events should generate a double resistant mutant every other generation. On the other hand, recombination is known to be localized to promoter regions, suggesting that these calculations may considerably overestimate intragenic levels of recombination (***Good, 2020***). To our knowledge, this is one of the first empirical examples of haplotype competition between beneficial mutations in wild plant populations. The mechanistic underpinnings of this haplotypic competition remains unresolved: there could be temporally fluctuating selection for one TSR allele or another (which is known to convey differential levels of protection to different chemical classes of ALS herbicides (***Patzoldt and Tranel, 2007***), negative epistasis between ALS574 and ALS653 that reduces their relative fitness advantage when found on the same background, or simply no fitness increase by stacking both alleles.

While we find that intra-chromosomal interactions have substantially impeded the selective trajectory of individual TSR alleles, we were also interested in the extent to which inter-chromosomal interactions may have facilitated herbicide resistance evolution. We find evidence that selection on Essex haplotypes containing ALS TSR mutations have likely been mediated by such interactions (Figure 4B-D). We find particularly extreme signed LD between TSR mutations and alleles on different chromosomes. LD between resistance alleles has been interpreted as epistatic selection, especially given the magnitude of selection from herbicides (***Gupta et al., 2021***), but LD between alleles that are not physically linked may also result from recent adaptive introgression. Recurrent gene-flow between resistant individuals is likely to result in genome-wide resistance alleles accumulating on the same background, with the shared population history of these resistance alleles overtime creating-cross chromosomal associations that could be interpreted as epistasis.

Alleles in windows on different chromosomes with the strongest evidence of interaction with ALS TSR mutations function in biological processes related to known organismal responses to ALS herbicides—compensating for reduced amino acid production via protein degradation and increased metabolism. These alleles may thus provide additional levels of tolerance on the large-effect TSR background, or may compensate for potential costs of these resistance mutations as has been seen for antibiotic resistance (***Craig MacLean et al., 2010***). Costs of the Trp-574-Leu mutation have been found in congeners of *A. tuberculatus* in terms of growth rate (***Tardif et al., 2006***), although there did not appear to be costs of Pro-653-Arg in *A. thaliana* (***Roux et al., 2004***), implying

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

that either of these fitness modifying mechanisms could be involved in the interactions we found here.

In conclusion, adaptation to herbicides and well-characterized target-site resistance mutations provide a powerful system for characterizing rapid and repeated evolution in wild plant populations, as well as the consequences of extreme selection on genome-wide patterns of diversity. Studies of resistance evolution have highlighted how extreme selection can modify life-history and plant mating systems (***Kuester et al., 2017***; ***Van Etten et al., 2020***) and vice versa (***Kreiner et al., 2018***), as well as the role of small- versus large-effect mutations (or monogenic versus polygenic adaptation) (***Kreiner et al.*** (***2020***); reviewed in ***Powles and Yu*** (***2010***); ***Délye*** (***2013***)), costs of adaptation under fluctuating environments (***Vila-Aiub et al., 2009***; ***Vila-Aiub, 2019***), and mutational repeatability (e.g. ***Menchari et al.*** (***2006***); ***Heap*** (***2014***)) (see ***Baucom*** (***2019***)). The work here contributes to this literature by characterizing extreme repeatability in origins of TSR resistance across a broad collection of agricultural populations as well as important causes—such as massive population growth over contemporary timescales that facilitates adaptation from readily available mutations—as well as consequences such parallelism. Namely, how selection varies considerably on identical mutations with distinct origins, in part as a result of differences in the timescale over which they have persisted, and in part resulting from inhibiting and facilitating allelic interactions within and across chromosomes. With increasing incidence of resistance and characterization of novel resistance alleles, population genomic approaches can continue to provide important insights into rapid evolutionary processes and factors that limit them, the importance of gene flow versus repeated origins in the spread of resistance across the range, and the timescale of evolution to contemporary environments.

Methods

Amaranthus tuberculatus sequence data

Sequencing and resequencing data were from a published study (***Kreiner et al., 2019***). Whole-genome Illumina sequencing data are available at European Nucleotide Archive (ENA), while the reference genome and its annotation are available on CoGe (reference ID = 54057). The analyses in this paper focus on herbicide resistance in 158 agricultural samples, collected from 8 fields with high *A. tuberculatus* densities across Missouri and Illinois in the Midwest United States (collected 2010), and from newly infested counties in Ontario, Canada, Walpole Island and Essex County (collected 2016). The 8 Midwestern populations had been surveyed for resistance to glyphosate herbicides in prior years (***Chatham et al., 2015***). 10 additional samples collected from natural populations in Ontario, Canada are also included, but only for tree-based inference. These samples have been recently analyzed with respect to the evolutionary origins of amplification of the glyphosate-targeted gene, 5-enolpyruvylshikimate-3-phosphate (***Kreiner et al., 2019***), as well as the polygenic architecture of glyphosate resistance (***Kreiner et al., 2020***).

SNP Calling & Phasing Genotypes

Filtered VCFs were from ***Kreiner et al.*** (***2019***) for all analyses. Briefly, freebayes-called SNPs were filtered based on missing data (>80% present), repeat content, allelic bias (>0.25 and <0.75), read paired status, and mapping quality (> Q30). Six individuals were removed due to excess missing data, leaving 152 agricultural and 10 natural samples for further analyses.

Known TSR mutations were assayed for presence/absence in our set of 162 A. tuberculatus individuals. At the time, that meant checking for known TSR mutations at 8 ALS amino acids (codons 122, 197, 205, 376, 377, 574, 653, 654), 3 PPO amino acids (***Rousonelos et al., 2012***; ***Giacomini et al., 2017***; ***Varanasi et al., 2018***), 1 PsbA (conferring resistance to photosystem II inhibitors) amino acid (***Lu et al., 2019***), and 3 EPSPS amino acids (***Perotti et al., 2019***). Briefly, to assay these mutations in our samples, we referred to the literature on previously verified TSR mutations in the genus *Amaranthus*, extracting the sequence surrounding a given focal TSR mutation, and BLAST (***Altschul***

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

et al., 1990) searched our reference genome to locate its position.

SHAPEIT2 (*Delaneau et al., 2013*), using joint inference from read-backed and population level information, was used to phase genotypes into haplotypes. Since phasing is very sensitive to data quality, we also applied a more stringent threshold of no more than 10% missing data for each SNP. SHAPEIT2 also requires a genetic map; with no genetic recombination map for *A. tuberculatus* yet available, we used LDhat to infer recombination rates across the genome in our samples (as in *Kreiner et al.* (*2019*)). Specifically, we used the interval function to estimate variable recombination rates within each of the 16 chromosomes of the pseudoassembly, using a precomputed lookup table for a θ of 0.01 for 192 chromosomes. We then converted rho estimates to genetic distance based recombination rates (100/$4N_eR$; N_e = 500,000), and used a monotonic spline to extrapolate genetic distance to each SNP in our VCF. We provided SHAPEIT an effective population size estimate of 500,000, inferred from previous demographic modelling in δaδi (*Kreiner et al., 2019*). As SHAPEIT only allows for biallelic SNPs, in order to retain the PPO210 deletion in phased output, we modified the alternate allele (a three base pair deletion) at the start position of the PPO210 codon to instead represent a simple SNP. This site otherwise retained the exact same genotype calls, allele frequencies, and other metadata.

Tree inference

Bootstrapped gene trees were inferred based on haplotypes within focal target-site genes (ALS and PPO), and 1 kb on either side around them. Using the phased data around these genes, we first converted each phased haplotype to FASTA format and then running clustal-ω2 across all samples, once for each gene, with 1,000 bootstraps. We then plotted mutational status for each focal TSR mutation (ALS574, ALS653, and PPO210) for each tip of both gene trees (Sup Figure 1).

We ran ARGweaver (*Rasmussen et al., 2014*; *Hubisz and Siepel, 2020*) on a region of 20,000 SNPs centered between the ALS and PPO genes on Scaffold 11. We used the settings -N (effective population size) 500,000 -r (recombination rate) 7e-9 -m (mutation rate) 1.8e-8 –ntimes (estimated timepoints) 50 –maxtime (max coalescent time) 100e3 -c (bp compression rate) 1 -n (MCMC samples) 500. We used an effective population size of 500,000, based on the best fitting demographic model previously inferred from this dataset with dadi (*Kreiner et al., 2019*). We then extracted the mostly likely ARG sample from the MCMC chain (sample 490/500), and the local trees corresponding to each of our three focal TSR mutations using arg-summarise. The arg-summarize function of ARGweaver was used to estimate the mean and 95% confidence intervals of the age of each mutational origin (based on clusters inferred from the most likely trees in the previous step) across the last 200 MCMC samples of the ARG. Since by default, arg-summarise –allele-age will infer the age of only the oldest allele under a scenario of multiple origins, we subsetted the dataset one mutational origin at a time (including all susceptible haplotypes) to obtain age estimates for all origins. For allele age estimates and coalescent tree-based tests of selection (see below), we only retained resistant haplotypes for a given focal mutation along with 20 haplotypes from natural populations that carry no TSR mutation (avoiding including multiple origins and inflating allele age, as well as the effects of linked selection from nearby resistance alleles on tree topology).

Coalescent tree-based tests for selection

RELATE (*Speidel et al., 2019*, *2021*), a scalable method for estimating tree-sequence across large genomic datasets, implements a tree-based test for detecting positive selection (*Griffiths and Tavaré, 1998*; *Speidel et al., 2019*). Under the standard coalescent model (i.e. assuming selective neutrality of mutations), the number of descendants in a particular lineage is exchangeable. Thus, one can compute the probability of some observed skew in the number of descendants using the hypergeometric distribution (*Griffiths and Tavaré, 1998*; *Speidel et al., 2019*). This approach gives us a *p*-value for this skew under the null (i.e., no selection). Since this statistic is simply based on order of coalescents, rather than branch lengths, it should be robust to misspecified (N_e) used to infer our ARG (*Speidel et al., 2019*). Since RELATE assumes an infinite sites model and thus is unsuit-

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

able for testing hypotheses about multiple origins, we performed our own implementation of this method for trees outputted from ARGweaver (***Rasmussen et al., 2014***).

Briefly, the statistic works as follows. Let f_N be the number of carriers of our focal mutation in the current day, N be the total present day sample size, and k_S be the number of susceptible lineages present when the mutation increases in count from 1 to 2. We sum each individual probability that a mutation spreads to at least a given frequency, from f_N to $N - k_S + 2$.

$$p_{R,denovo} = \sum_{f=f_N}^{N-k_S+2} \frac{(f-1)\binom{N-f-1}{k_S-3}}{\binom{N-1}{k_S-1}} \tag{1}$$

The null hypothesis, that allele frequency change occurred under drift, is rejected when this one-sided *p*-value is sufficiently small (i.e. $p < 0.05$), implying selection has governed the spread of this mutation since it first arose.

We also modified this statistic to test for selection on more recent timescales, and thus the scenario of adaptation from standing genetic variation. Here, we need to define k_R, the number of resistant lineages at some time (t) before the present day, in addition to $k_S(t)$.

$$p_{R,sgv} = \sum_{f=f_N}^{N-k_S+2} \frac{\binom{f-1}{k_R(t)-1}\binom{N-f-1}{k_S(t)-1}}{\binom{N-1}{k_R(t)+k_S(t)-1}} \tag{2}$$

The null hypothesis that the frequency change (between the current day and some time in the past more recent than when the mutation first arose (t)) happened under random drift (and hence no selective pressures) is rejected if this *p*-value is sufficiently small.

N_e estimation through time and selection coefficients

We used RELATE 1.1.6 (***Speidel et al., 2021***) to estimate tree sequence from distinct recombinational units across the genome from our phased dataset. RELATE requires polarized ancestral allele calls, such that alternate alleles represent the derived state. To do so, we performed a multiple alignment of our *A. tuberculatus* genome to *A. palmeri* (***Montgomery et al., 2020***) using lastz (***Harris, 2007***), retained the best orthologous chain from the alignment, and extracted variant sites. We modified the *A. tuberculatus* reference genome with the derived allele states from our multiple alignment, using this modified reference to polarize allele calls. On each chromosome, we then ran RelateParallel.sh –mode All, using the output from all chromosomes to first estimate mutation rate (RelateMutationRate –mode Avg). We then reestimated branch lengths with this updated mutation rate (ReEstimateBranchLengths), and lastly estimated population size through time (EstimatePopulationSize.sh). Population sizes were estimated from 0 years ago to 10,000,000 years ago, in epoch timesteps of $10^{0.25}$ years, to obtain particularly fine-scale estimates in the recent past.

We estimated selection coefficients using (***Stern et al., 2019***) for the origin with the strongest signatures of selection, ALS653 #7. Similar to our tree-based estimates of selection, we subsetted each tree to include just resistant haplotypes of our focal mutational origin and a reference set of 20 haplotypes susceptible for all three common TSR alleles. After subsetting, we re-estimated branch lengths and sampled branches corresponding to the locus of interest. Finally, we performed Bayesian inference of the selection coefficient of our mutational origin, accounting for changing population sizes through time as estimated in RELATE. For these tests, we used the Ontario allele frequency (0.29) for the -popFreq parameter, and ran the analysis twice, once for epoch 0-30, and once for 0-10.

Selection Scans & LD-based analyses

The phased data used as input for ARGweaver was also used to extract selective sweep summary statistics in selscan (***Szpiech and Hernandez, 2014***). In selscan, we estimated both XPEHH (***Sabeti***

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

694 ***et al., 2007***), in this case, the difference in the integrated extended haplotype homozygosity be-
695 tween resistant and susceptible haplotypes, and mean pairwise difference estimates. For both of
696 these statistics, we provided LD-based recombination maps, inferred from LDhat, as described
697 above. Because some individuals in Essex always carried at least one resistant ALS haplotype
698 through either mutations at ALS574 or ALS653, to compare patterns of selection associated with re-
699 sistance and susceptibility, these statistics were calculated at the haplotype, rather than individual
700 level.

701 We used plink (v1.90b3.46) (***Purcell et al., 2007***) to calculate both r^2 between each focal TSR
702 mutations and missense mutations on the same chromosome, and to estimate signed LD (r) both
703 between TSR mutations, and with all other bi-allelic missense SNPs across the genome. We per-
704 formed these calculations with respect to a given TSR mutation by using the –ld-snp options to
705 specify a focal mutation. To visualize patterns of signed LD between TSR mutations and other mis-
706 sense SNPs, we split the genome into non-overlapping 10 kb windows and calculated the average
707 LD among all SNPs in each window. All LD calculations were polarized by rarity (e.g. minor alleles
708 segregating on the same haplotypes were regarded as being in positive LD). In Essex, despite be-
709 ing considerably common, both ALS 574 and ALS 653 had a frequency less than 50%, so LD values
710 between all missense alleles, and both these focal TSR mutations are directly comparable.

711 To test whether the top 2% of 1 Mb windows of missense SNPs with particularly low or high
712 signed LD with ALS TSR mutations (either ALS653 or ALS574) was significantly different from the
713 null expectation, we used a permutation approach whereby we randomly shuffled the assignment
714 of the focal ALS TSR mutation between all individuals and calculated mean LD (with respect to the
715 permuted TSR mutations) in the window of interest. We then repeated this permutation 1,000
716 times to generate a null distribution for comparison to the real average signed LD value of each
717 region. This permutation test explicitly evaluates whether a TSR mutation and missense mutations
718 in a focal window are more likely to be found together in the specific set of individuals containing
719 the focal TSR mutation than any other set of individuals of the same size. Thus, this test is robust to
720 variance in missing data. The proportion of permuted observations with a mean absolute signed
721 LD exceeding the observed signed LD was taken as the two-tailed p-value for each windowed test.
722 Lastly, we found the intersect of these windows with the closest gene according to our genome
723 annotation, and found their *A. thaliana* orthologs (***Emms and Kelly, 2015***). We used the set of *A.*
724 *thaliana* orthologs found across all 13 significantly enriched 1 Mb windows (after a 5% FDR correc-
725 tion) in a Gene Ontology (GO) Enrichment analysis for biological processes.

726 Acknowledgments

727 We gratefully acknowledge discussion and feedback on the manuscript from Matt Osmond, Aneil
728 Agrawal, Rob Ness, and Sally Otto. Thanks to Leo Speidel and Melissa Hubisz for input on RELATE
729 and ARGweaver implementation, and Peter Sikkema and lab with their support on sampling On-
730 tario populations. J.M.K. was supported by the Rosemary Grant Advanced Award from the Society
731 for the Study of Evolutionary Biology and NSERC PGS-D, D.W. was supported by the Max Planck
732 Society, J.R.S. and S.I.W. were supported by Discovery Grants from NSERC Canada, and S.I.W. was
733 additionally supported by a Canada Research Chair in Population Genomics.

734 References

735 **Altschul SF**, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct;
736 215(3):403–410.

737 **Barrick JE**, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF. Genome evolution and adaptation
738 in a long-term experiment with Escherichia coli. Nature. 2009 Oct; 461(7268):1243–1247.

739 **Baucom RS**. Evolutionary and ecological insights from herbicide-resistant weeds: what have we learned about
740 plant adaptation, and what is left to uncover? New Phytol. 2019 Jul; 223(1):68–82.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Bell G. Evolutionary rescue and the limits of adaptation. Philos Trans R Soc Lond B Biol Sci. 2013 Jan; 368(1610):20120080.

Brown HM. Mode of action, crop selectivity, and soil relations of the sulfonylurea herbicides. Pestic Sci. 1990 May; 29(3):263–281.

Busi R, Gaines TA, Walsh MJ, Powles SB. Understanding the potential for resistance evolution to the new herbicide pyroxasulfone: field selection at high doses versus recurrent selection at low doses: Resistance evolution by high- and low-dose selection. Weed Res. 2012 Dec; 52(6):489–499.

Busi R, Powles SB. Evolution of glyphosate resistance in a Lolium rigidum population by glyphosate selection at sublethal doses. Heredity. 2009 Oct; 103(4):318–325.

Busi R, Neve P, Powles S. Evolved polygenic herbicide resistance in Lolium rigidum by low-dose herbicide selection within standing genetic variation. Evol Appl. 2013 Feb; 6(2):231–242.

Castellano D, Coronado-Zamora M, Campos JL, Barbadilla A, Eyre-Walker A. Adaptive Evolution Is Substantially Impeded by Hill–Robertson Interference in Drosophila. Mol Biol Evol. 2015 Oct; 33(2):442–455.

Charlesworth B. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. Nat Rev Genet. 2009 Mar; 10(3):195–205.

Chatham LA, Wu C, Riggins CW, Hager AG, Young BG, Roskamp GK, Tranel PJ. EPSPS Gene Amplification is Present in the Majority of Glyphosate-Resistant Illinois Waterhemp (Amaranthus tuberculatus) Populations. Weed Technol. 2015 Mar; 29(1):48–55.

Chen L, Su ZZ, Huang L, Xia FN, Qi H, Xie LJ, Xiao S, Chen QF. The AMP-Activated Protein Kinase KIN10 Is Involved in the Regulation of Autophagy in Arabidopsis. Front Plant Sci. 2017 Jul; 8:1201.

Comai L, Stalker D. Mechanism of action of herbicides and their molecular manipulation. . 1986; 3:166–195.

ffrench Constant RH, Daborn PJ, Goff GL. The genetics and genomics of insecticide resistance. Trends Genet. 2004; 20(3):163–170.

Costea M, Weaver SE, Tardif FJ. The Biology of Invasive Alien Plants in Canada. 3. Amaranthus tuberculatus (Moq.) Sauer var. rudis (Sauer) Costea & Tardif. Can J Plant Sci. 2005 Apr; 85(2):507–522.

Craig MacLean R, Hall AR, Perron GG, Buckling A. The population genetics of antibiotic resistance: integrating molecular mechanisms and treatment contexts. Nat Rev Genet. 2010 Jun; 11(6):405–414.

Dayan FE, Barker A, Tranel PJ. Origins and structure of chloroplastic and mitochondrial plant protoporphyrinogen oxidases: implications for the evolution of herbicide resistance. Pest Manag Sci. 2018 Oct; 74(10):2226–2234.

Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods. 2013 Jan; 10(1):5–6.

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

Eldholm V, Norheim G, von der Lippe B, Kinander W, Dahle UR, Caugant DA, Mannsåker T, Mengshoel AT, Dyrhol-Riise AM, Balloux F. Evolution of extensively drug-resistant Mycobacterium tuberculosis from a susceptible ancestor in a single patient. Genome Biol. 2014 Nov; 15(11):490.

Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015 Aug; 16:157.

Faus I, Zabalza A, Santiago J, Nebauer SG, Royuela M, Serrano R, Gadea J. Protein kinase GCN2 mediates responses to glyphosate in Arabidopsis. BMC Plant Biol. 2015 Jan; 15:14.

Feder AF, Kline C, Polacino P, Cottrell M, Kashuba ADM, Keele BF, Hu SL, Petrov DA, Pennings PS, Ambrose Z. A spatio-temporal assessment of simian/human immunodeficiency virus (SHIV) evolution reveals a highly dynamic process within the host. PLoS Pathog. 2017 May; 13(5):e1006358.

Feder AF, Pennings PS, Hermisson J, Petrov DA. Evolutionary Dynamics in Structured Populations Under Strong Population Genetic Forces. G3. 2019 Oct; 9(10):3395–3407.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Feder AF, Rhee SY, Holmes SP, Shafer RW, Petrov DA, Pennings PS. More effective drugs lead to harder selective sweeps in the evolution of drug resistance in HIV-1. Elife. 2016 Feb; 5.

Flood PJ, van Heerwaarden J, Becker F, de Snoo CB, Harbinson J, Aarts MGM, Whole-Genome Hitchhiking on an Organelle Mutation; 2016.

Fujiki Y, Yoshikawa Y, Sato T, Inada N, Ito M, Nishida I, Watanabe A, Dark-inducible genes from Arabidopsis thaliana are associated with leaf senescence and repressed by sugars; 2001.

Gerrish PJ, Lenski RE. The fate of competing beneficial mutations in an asexual population. Genetica. 1998; 102-103(1-6):127–144.

Giacomini DA, Umphres AM, Nie H, Mueller TC, Steckel LE, Young BG, Scott RC, Tranel PJ. Two new PPX2 mutations associated with resistance to PPO-inhibiting herbicides in Amaranthus palmeri: PPX2 mutations. Pest Manag Sci. 2017 Aug; 73(8):1559–1563.

Good BH. Linkage disequilibrium between rare mutations; 2020.

Griffiths RC, Marjoram P. Ancestral inference from samples of DNA sequences with recombination. J Comput Biol. 1996; 3(4):479–502.

Griffiths RC, Tavaré S. The age of a mutation in a general coalescent tree. Communications in Statistics Stochastic Models. 1998 Jan; 14(1-2):273–295.

Griffiths RC, Marjoram P, An Ancestral Recombination Graph; 1997.

Gupta S, Harkess A, Soble A, Van Etten M, Leebens-Mack J, Baucom RS. Inter-chromosomal linkage disequilibrium and linked fitness cost loci influence the evolution of nontarget site herbicide resistance in an agricultural weed; 2021.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 2009 Oct; 5(10):e1000695.

Harris RS. Improved pairwise alignmnet of genomic DNA. . 2007 Sep; .

Heap I. Herbicide Resistant Weeds. In: Pimentel D, Peshin R, editors. *Integrated Pest Management* Springer Netherlands; 2014.p. 281–301.

Hedrick PW. Genetic polymorphism in a temporally varying environment: effects of delayed germination or diapause. Heredity. 1995 Aug; 75(2):164–170.

Hegreness M, Shoresh N, Hartl D, Kishony R. An equivalence principle for the incorporation of favorable mutations in asexual populations. Science. 2006 Mar; 311(5767):1615–1617.

Hejase HA, Salman-Minkov A, Campagna L, Hubisz MJ, Lovette IJ, Gronau I, Siepel A. Genomic islands of differentiation in a rapid avian radiation have been driven by recent selective sweeps. Proc Natl Acad Sci U S A. 2020 Dec; 117(48):30554–30565.

Hill WG, Robertson A, The effect of linkage on limits to artificial selection; 1966.

Holst N, Rasmussen IA, Bastiaans L. Field weed population dynamics: a review of model approaches and applications. Weed Res. 2007 Feb; 47(1):1–14.

Hubisz M, Siepel A. Inference of Ancestral Recombination Graphs Using ARGweaver. Methods Mol Biol. 2020; 2090:231–266.

Hudson RR. Properties of a neutral allele model with intragenic recombination. Theor Popul Biol. 1983 Apr; 23(2):183–201.

Karasov T, Messer PW, Petrov DA. Evidence that adaptation in Drosophila is not limited by mutation at single sites. PLoS Genet. 2010 Jun; 6(6):e1000924.

Kreiner JM, Giacomini DA, Bemm F, Waithaka B, Regalado J, Lanz C, Hildebrandt J, Sikkema PH, Tranel PJ, Weigel D, Stinchcombe JR, Wright SI. Multiple modes of convergent adaptation in the spread of glyphosate-resistant Amaranthus tuberculatus. Proc Natl Acad Sci U S A. 2019 Oct; 116(42):21076–21084.

Kreiner JM, Tranel PJ, Weigel D, Stinchcombe JR, Wright SI. The genetic architecture and genomic context of glyphosate resistance; 2020.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Kreiner JM, Stinchcombe JR, Wright SI. Population Genomics of Herbicide Resistance: Adaptation via Evolutionary Rescue. Annu Rev Plant Biol. 2018 Apr; 69:611–635.

Kuester A, Fall E, Chang SM, Baucom RS. Shifts in outcrossing rates and changes to floral traits are associated with the evolution of herbicide resistance in the common morning glory. Ecol Lett. 2017 Jan; 20(1):41–49.

Küpper A, Manmathan HK, Giacomini D, Patterson EL, McCloskey WB, Gaines TA. Population Genetic Structure in Glyphosate-Resistant and -Susceptible Palmer Amaranth (Amaranthus palmeri) Populations Using Genotyping-by-sequencing (GBS). Front Plant Sci. 2018 Jan; 9:29.

Lenormand T, Harmand N, Gallet R, Cost of resistance: an unreasonably expensive concept; 2018.

Leslie T, Baucom RS. 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. 2014 Oct; 4(10):2035–2047.

Lu H, Yu Q, Han H, Owen MJ, Powles SB. A novel psbA mutation (Phe274-Val) confers resistance to PSII herbicides in wild radish (Raphanus raphanistrum). Pest Manag Sci. 2019 Jan; 75(1):144–151.

Marshall R, Hull R, Moss SR. Target site resistance to ALS inhibiting herbicides in Papaver rhoeas and Stellaria media biotypes from the UK. Weed Res. 2010 Dec; 50(6):621–630.

Massa D, Krenz B, Gerhards R. Target-site resistance to ALS-inhibiting herbicides in Apera spica-venti populations is conferred by documented and previously unknown mutations. Weed Res. 2011; .

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

McNaughton KE, Lee EA, Tardif FJ. Mutations in the ALS gene conferring resistance to group II herbicides in redroot pigweed (Amaranthus retroflexus) and green pigweed (A. powellii). In: *Weed Sci. Soc. Am. Abstr*, vol. 41; 2001. p. 97.

Menchari Y, Camilleri C, Michel S, Brunel D, Dessaint F, Le Corre V, Délye C. Weed response to herbicides: regional-scale distribution of herbicide resistance alleles in the grass weed Alopecurus myosuroides. New Phytol. 2006; 171(4):861–874.

Molin WT, Wright AA, Lawton-Rauh A, Saski CA. The unique genomic landscape surrounding the EPSPS gene in glyphosate resistant Amaranthus palmeri: a repetitive path to resistance. BMC Genomics. 2017 Jan; 18(1):91.

Montgomery JS, Giacomini D, Waithaka B, Lanz C, Murphy BP, Campe R, Lerchl J, Landes A, Gatzmann F, Janssen A, Antonise R, Patterson E, Weigel D, Tranel PJ. Draft Genomes of Amaranthus tuberculatus, Amaranthus hybridus, and Amaranthus palmeri. Genome Biol Evol. 2020 Nov; 12(11):1988–1993.

Murphy BP, Tranel PJ. Target-Site Mutations Conferring Herbicide Resistance. Plants. 2019 Sep; 8(10).

Nakka S, Thompson CR, Peterson DE, Jugulam M. Target Site–Based and Non–Target Site Based Resistance to ALS Inhibitors in Palmer Amaranth (Amaranthus palmeri). Weed Sci. 2017 Nov; 65(6):681–689.

Naylor REL, WEEDS | Weed Biology; 2003.

Neve P, Busi R, Renton M, Vila-Aiub MM. Expanding the eco-evolutionary context of herbicide resistance research. Pest Manag Sci. 2014 Sep; 70(9):1385–1393.

Neve P, Powles S. Recurrent selection with reduced herbicide rates results in the rapid evolution of herbicide resistance in Lolium rigidum. Theor Appl Genet. 2005 Apr; 110(6):1154–1166.

Neve P, Vila-Aiub M, Roux F. Evolutionary-thinking in agricultural weed management. New Phytol. 2009; 184(4):783–793.

Orcaray L, Igal M, Zabalza A, Royuela M. Role of exogenously supplied ferulic and p-coumaric acids in mimicking the mode of action of acetolactate synthase inhibiting herbicides. J Agric Food Chem. 2011 Sep; 59(18):10162–10168.

Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, Shaw RG, Weigel D, Lynch M. The rate and molecular spectrum of spontaneous mutations in Arabidopsis thaliana. Science. 2010 Jan; 327(5961):92–94.

Panozzo S, Scarabel L, Tranel PJ, Sattin M. Target-site resistance to ALS inhibitors in the polyploid species Echinochloa crus-galli. Pestic Biochem Physiol. 2013 Feb; 105(2):93–101.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Patzoldt WL, Tranel PJ. Multiple ALS Mutations Confer Herbicide Resistance in Waterhemp (Amaranthus tuberculatus). Weed Sci. 2007 Oct; 55(5):421–428.

Pennings PS, Kryazhimskiy S, Wakeley J. Loss and recovery of genetic diversity in adapting populations of HIV. PLoS Genet. 2014 Jan; 10(1):e1004000.

Perotti VE, Larran AS, Palmieri VE, Martinatto AK, Alvarez CE, Tuesca D, Permingeat HR. A novel triple amino acid substitution in the EPSPS found in a high-level glyphosate-resistant Amaranthus hybridus population from Argentina. Pest Manag Sci. 2019 May; 75(5):1242–1251.

Peterson MA, Collavo A, Ovejero R, Shivrain V, Walsh MJ. The challenge of herbicide resistance around the world: a current summary. Pest Manag Sci. 2018 Oct; 74(10):2246–2259.

Petit C, Duhieu B, Boucansaud K, Délye C. Complex genetic control of non-target-site-based resistance to herbicides inhibiting acetyl-coenzyme A carboxylase and acetolactate-synthase in Alopecurus myosuroides Huds. Plant Sci. 2010; 178(6):501–509.

Powles SB, Yu Q. Evolution in action: plants resistant to herbicides. Annu Rev Plant Biol. 2010; 61:317–347.

Preston C. Inheritance and linkage of metabolism-based herbicide cross-resistance in rigid ryegrass (Lolium rigidum). wees. 2003 Jan; 51(1):4–12.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep; 81(3):559–575.

Ralph P, Coop G. Parallel adaptation: one or many waves of advance of an advantageous allele? Genetics. 2010 Oct; 186(2):647–668.

Rasmussen MD, Hubisz MJ, Gronau I, Siepel A. Genome-wide inference of ancestral recombination graphs. PLoS Genet. 2014 May; 10(5):e1004342.

Rousonelos SL, Lee RM, Moreira MS, VanGessel MJ, Tranel PJ, Characterization of a Common Ragweed (Ambrosia artemisiifolia) Population Resistant to ALS- and PPO-Inhibiting Herbicides; 2012.

Roux F, Gasquez J, Reboud X. The dominance of the herbicide resistance cost in several Arabidopsis thaliana mutant lines. Genetics. 2004 Jan; 166(1):449–460.

Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, Schaffner SF, Lander ES, International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, et al. Genome-wide detection and characterization of positive selection in human populations. Nature. 2007 Oct; 449(7164):913–918.

Salas RA, Burgos NR, Tranel PJ, Singh S, Glasgow L, Scott RC, Nichols RL. Resistance to PPO-inhibiting herbicide in Palmer amaranth from Arkansas. Pest Manag Sci. 2016 May; 72(5):864–869.

Shoup DE, Al-Khatib K, Peterson DE. Common waterhemp (Amaranthus rudis) resistance to protoporphyrinogen oxidase-inhibiting herbicides. Weed Sci. 2003 Apr; 51(2):145–150.

Singh S, Singh V, Salas-Perez RA, Bagavathiannan MV, Lawton-Rauh A, Roma-Burgos N. Target-site mutation accumulation among ALS inhibitor-resistant Palmer amaranth. Pest Manag Sci. 2018 Dec; 8:350.

Speidel L, Cassidy L, Davies RW, Hellenthal G, Skoglund P, Myers SR. Inferring population histories for ancient genomes using genome-wide genealogies; 2021.

Speidel L, Forest M, Shi S, Myers SR. A method for genome-wide genealogy estimation for thousands of samples. Nat Genet. 2019 Sep; 51(9):1321–1329.

Stern AJ, Wilton PR, Nielsen R. An approximate full-likelihood method for inferring selection and allele frequency trajectories from DNA sequence data. PLoS Genet. 2019 Sep; 15(9):e1008384.

Sun G, Luo T, Yang C, Dong X, Li J, Zhu Y, Zheng H, Tian W, Wang S, Barry CE 3rd, Mei J, Gao Q. Dynamic population changes in Mycobacterium tuberculosis during acquisition and fixation of drug resistance in patients. J Infect Dis. 2012 Dec; 206(11):1724–1733.

Szpiech ZA, Hernandez RD. selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. Mol Biol Evol. 2014 Oct; 31(10):2824–2827.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

928 **Tardif FJ**, Rajcan I, Costea M. A mutation in the herbicide target site acetohydroxyacid synthase produces
929 morphological and structural alterations and reduces fitness in Amaranthus powellii. New Phytol. 2006;
930 169(2):251–264.

931 **Tranel PJ**, Wright TR. Resistance of weeds to ALS-inhibiting herbicides: what have we learned? Weed Sci. 2002
932 Nov; 50(6):700–712.

933 **Trenkamp S**, Eckes P, Busch M, Fernie AR. Temporally resolved GC-MS-based metabolic profiling of herbicide
934 treated plants treated reveals that changes in polar primary metabolites alone can distinguish herbicides of
935 differing mode of action. Metabolomics. 2009 Sep; 5(3):277–291.

936 **USDA-NASS**, Agricultural Chemical Usage: Field Crops Summary; 2012.

937 **Van Etten M**, Lee KM, Chang SM, Baucom RS. Parallel and nonparallel genomic responses contribute to herbi-
938 cide resistance in Ipomoea purpurea, a common agricultural weed; 2019.

939 **Van Etten M**, Soble A, Baucom RS. Variable inbreeding depression may explain associations between the
940 mating system and herbicide resistance; 2020.

941 **Varanasi VK**, Brabham C, Norsworthy JK, Nie H, Young BG, Houston M, Barber T, Scott RC. A Statewide Survey
942 of PPO-Inhibitor Resistance and the Prevalent Target-Site Mechanisms in Palmer amaranth (Amaranthus
943 palmeri) Accessions from Arkansas. Weed Sci. 2018 Mar; 66(2):149–158.

944 **Vila-Aiub MM**. Fitness of Herbicide-Resistant Weeds: Current Knowledge and Implications for Management.
945 Plants. 2019 Nov; 8(11).

946 **Vila-Aiub MM**, Neve P, Powles SB. Fitness costs associated with evolved herbicide resistance alleles in plants.
947 New Phytol. 2009 Dec; 184(4):751–767.

948 **de Visser JAGM**, Rozen DE. Limits to adaptation in asexual populations. J Evol Biol. 2005 Jul; 18(4):779–788.

949 **Vitalis R**, Glémin S, Olivieri I. When genes go to sleep: the population genetic consequences of seed dormancy
950 and monocarpic perenniality. Am Nat. 2004 Feb; 163(2):295–311.

951 **Weng ML**, Becker C, Hildebrandt J, Neumann M, Rutter MT, Shaw RG, Weigel D, Fenster CB. Fine-Grained
952 Analysis of Spontaneous Mutation Spectrum and Frequency in Arabidopsis thaliana. Genetics. 2019 Feb;
953 211(2):703–714.

954 **Whaley CM**, Wilson HP, Westwood JH. Characterization of a new ALS-inhibitor resistance mutation from the
955 ALS gene of smooth pigweed (Amaranthus hybridus). In: *Weed Sci. Soc. Amer. Abstr*, vol. 44; 2004. p. 161.

956 **Whitcomb CE**. An introduction to ALS-inhibiting herbicides. Toxicol Ind Health. 1999 Jan; 15(1-2):231–239.

957 **Williams KA**, Pennings P. Drug Resistance Evolution in HIV in the Late 1990s: Hard Sweeps, Soft Sweeps, Clonal
958 Interference and the Accumulation of Drug Resistance Mutations. G3. 2020 Apr; 10(4):1213–1223.

959 **Zhao L**, Deng L, Zhang Q, Jing X, Ma M, Yi B, Wen J, Ma C, Tu J, Fu T, Shen J. Autophagy contributes to sulfonylurea
960 herbicide tolerance via GCN2-independent regulation of amino acid homeostasis. Autophagy. 2018 Jan;
961 14(4):702–714.

962 **Zulet A**, Gil-Monreal M, Villamor JG, Zabalza A, van der Hoorn RAL, Royuela M. Proteolytic pathways induced
963 by herbicides that inhibit amino acid biosynthesis. PLoS One. 2013 Sep; 8(9):e73847.

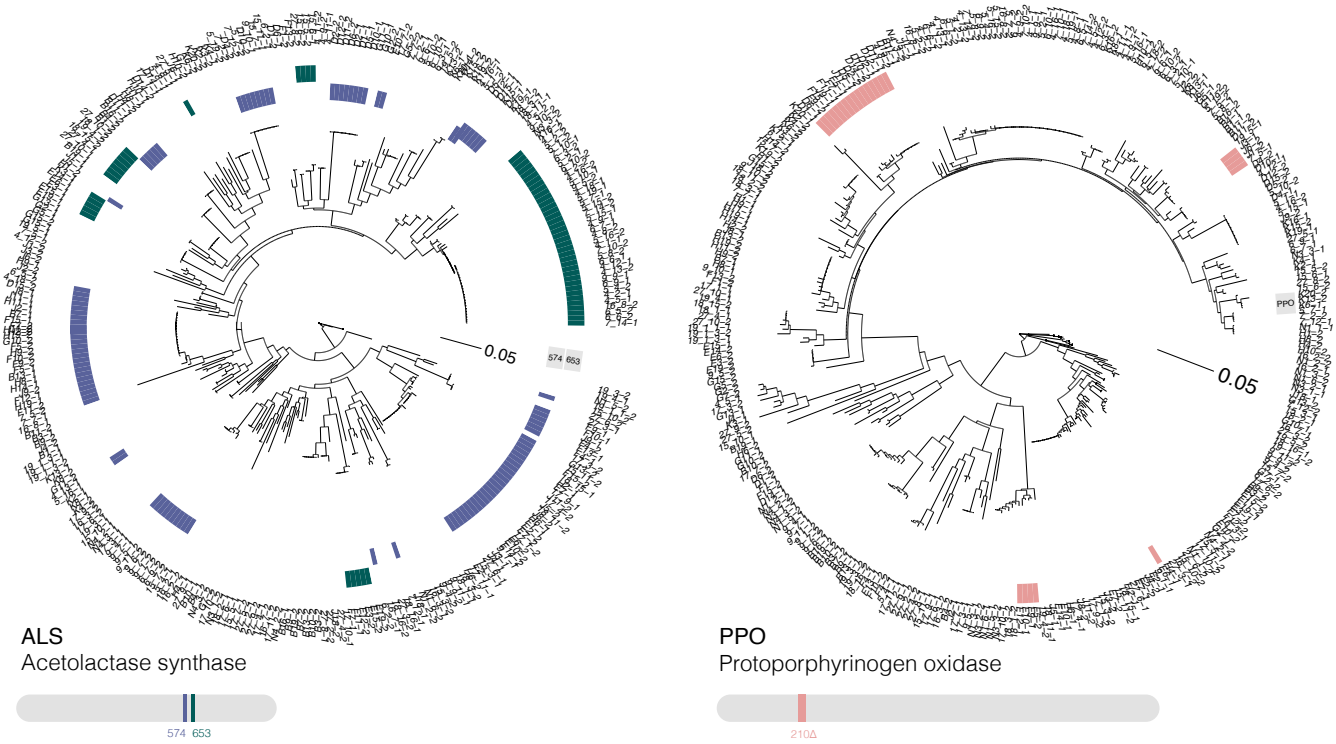
bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Supplementary Material

Sup Table 1. Tree-based coalescent test for selection under two scenarios; selection on *de novo* origins (since the mutation first arose) and selection over recent timescales (consistent with selection on a standing genetic variant). All bolded significant p-values remain significant after a 5% FDR correction.

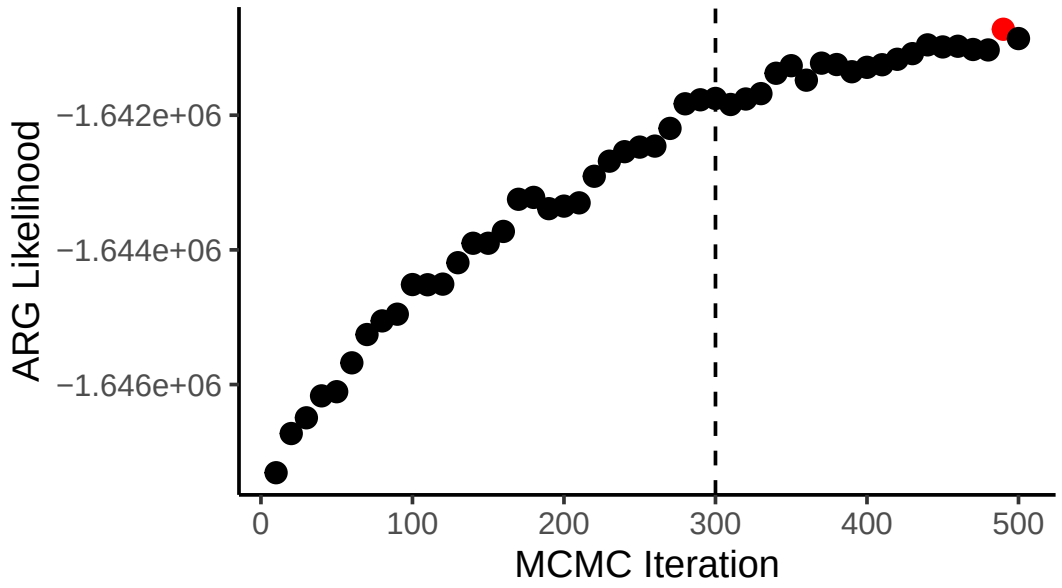
TSR position	Mutational Origin #	Selection *de novo*, p-value	Selection on SGV, p-value
ALS574	1	**9.81E-06**	0.2449
ALS574	2	**0.000885252**	**-**
ALS574	3	**0.005650493**	0.9800
ALS574	4	0.1401669	**0.0013**
ALS574	5	0.6623966	0.3289
ALS574	6	0.6397279	0.3668
ALS653	7	**5.58E-06**	**2.67E-14**
ALS653	8	0.7168271	0.0765
PPO210	9	0.5886909	0.5470
PPO210	10	0.2841429	0.2703
PPO210	11	0.5886909	0.7124

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

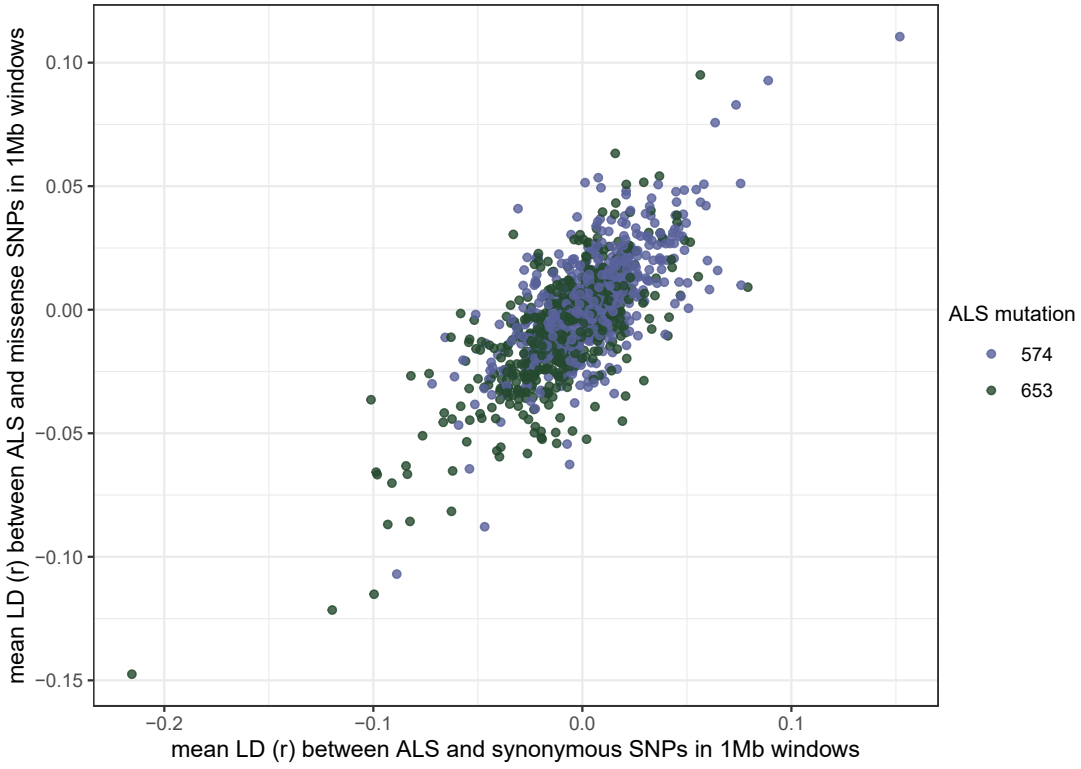


Sup Figure 1. Bootstrapped gene trees of ALS (3 kb) and PPO (10 kb) (CDS +1 kb on either side) alongside TSR mutations across all 162 individuals. Coloured grids indicate a haplotype harbouring a focal resistance SNP, with the SNP legend inserted within the grid (ALS574 mutations in green, ALS653 mutations in purple, PPO210 deletions in pink). Multiple origins are apparent for all resistance mutations.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.10.443516; this version posted May 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.



Sup Figure 2. Ancestral recombination graph likelihood by MCMC iteration in ARGweaver for a 20,000 SNP region centered around ALS and PPO. Red dot highlights the ARG that maximizes the likelihood of the model fit to our data, with which tree-based inferences were conducted. All iterations after 300 were used for estimating the mean and 95% confidence intervals of allelic ages of mutational origins.



Sup Figure 3. Correlation of signed LD (*r*) between two target site resistance mutations, and synonymous or missense mutations across the genome. Each point represents the mean r between all synonymous or missense mutations and an ALS mutation in a non-overlapping 1 Mb window.