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 agricultural fields are well-characterized for herbicide resistance. However, we still lack an understanding of their evolutionary history: the extent of parallelism in the origins of target-site resistance (TSR), how long these mutations persist, how quickly they spread, and allelic interactions that mediate their selective advantage. We addressed these questions with genomic data from 18 agricultural populations of common waterhemp (*Amaranthus tuberculatus*), which we show to have undergone a massive expansion over the past century with a contemporary effective population size estimate of 8×10^7. We found variation at seven characterized TSR loci, two of which had multiple amino acid substitutions, and three of which were common. These three common resistance variants show parallelism in their mutational origins, with gene flow having shaped their distribution across the landscape. Allele age estimates supported a strong role of adaptation from *de novo* mutations, with a median allele age of 30 suggesting that most resistance alleles arose soon after the onset of herbicide use. However, resistant lineages varied in both their age and evidence for selection over two different timescales, implying considerable heterogeneity in the forces that govern their persistence. The evolutionary history of TSR has also been shaped by both intra- and inter-locus allelic interactions. We report a signal of extended haplotype competition between two common TSR alleles, and extreme linkage with genome-wide alleles with known functions in resistance adaptation. Together, this work reveals a remarkable example of spatial parallel evolution in a metapopulation, with important implications for the management of herbicide resistance.

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—over 500 across the globe—based on just the occurrence of resistance to different herbicide modes-of-action across different species (***Heap, 2021***) 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 in 1986, which was only five years after their initial introduction (***Comai and Stalker, 1986***; ***Heap, 2021***; ***Whitcomb,***

1999). These numbers are likely a vast underestimate of the repeatability of herbicide resistance evolution. For ALS herbicides, for example, non-synonymous substitutions at eight distinct codons confer resistance, with most of them found in multiple species (***Tranel and Wright, 2002***), and with multiple independent causal mutations often occurring in the same population (***Heap, 2021***; ***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 in plants, it 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 (***Kuester et al., 2017***), 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 (***Kreiner et al., 2021***; ***Van Etten et al., 2019***). 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 and informing integrative management strategies in weed populations.

In contrast to most of the selective sweep literature coming from within-host studies of drug resistance in HIV (e.g. ***Feder et al.*** (***2016***); ***Pennings et al.*** (***2014***))—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 and changing selection regimes through rotations of both focal crops and herbicide applications (***Holst et al., 2007***; ***Naylor, 2003***; ***Neve et al., 2009***). Consequently, persistent agricultural weed populations likely comprise a collection of resistant haplotypes that have arisen and dispersed across the landscape, following a model of spatial parallel mutation in an interconnected network of populations (***Ralph and Coop, 2010***). Recent population genomic evidence supports this prediction for a subset of newly problematic glyphosate-resistant agricultural populations of *Amaranthus tuberculatus* in Ontario, Canada, where both multiple origins and long-distance dispersal contributed to the spread of glyphosate resistance (***Kreiner et al., 2019***).

Amaranthus tuberculatus, or common waterhemp, is a major challenge for agricultural practices in the Midwestern US, and is among the most problematic weeds worldwide in terms of its capacity for evolving resistance to multiple herbicides (***Tranel, 2021***). Conforming to classic hypotheses about successful weeds (***Baker, 1974***), the species has large census population sizes, a widespread distribution, and considerable seedbanks (***Costea et al., 2005***). The dioecious, wind-pollinated *A. tuberculatus* additionally offers an obligately outcrossing mating system, providing more independent backgrounds on which new mutations arise (***Costea et al., 2005***; ***Kreiner et al., 2018***) and an effective dispersal system (***Liu et al., 2012***). The species has not always been troublesome—the plant is native to North America, where it likely has grown in riparian habitats long before the advent of modern agriculture (***Sauer, 1957***). While it persists in these habitats, in the past 100 years or so, *A. tuberculatus* has become strongly associated with agricultural fields (***Costea et al., 2005***; ***Tranel and Trucco, 2009***). Over the past three decades, *A. tuberculatus* has evolved resistance to seven chemical modes-of-action (***Heap, 2021***; ***Tranel, 2021***), including resistance to both ALS-inhibiting herbicides and protoporphyrinogen oxidase (PPO)-inhibiting herbicides (***Shoup et al., 2003***).

ALS-inhibiting herbicides have been among the most popular mode-of-action for weed control

in crops since their introduction in the 1980s (***Brown, 1990***), and are widely used in both corn and soy production systems. They were rapidly adopted due to their application rates being an order of magnitude lower than previous herbicides and thus increased affordability, along with low toxicity and broad-spectrum weed control (***Mazur and Falco, 1989***), but quickly became notorious for their ability to select for resistant weeds (***Tranel and Wright, 2002***). Use of ALS herbicides thus decreased in the 1990s, coinciding with the widespread adoption of Round-up ready cropping systems, of which glyphosate herbicides are an essential component (***Green, 2007***). While historically not as popular as ALS-inhibiting herbicides, PPO-inhibiting herbicides have been used for nearly 50 years for the control of dicotyledonous (broadleaf) weeds, at its peak in the early 1990s representing 10% of annual applications in the USA but dropping to 1.5% by 2006 (***USDA-NASS, 2012***). However, PPO-inhibiting herbicides have since seen a resurgence for control of weeds that have evolved resistance to heavily-used herbicides such as glyphosate and ALS inhibitors (***Tranel and Wright, 2002***; ***Dayan et al., 2018***; ***Zhao et al., 2020***; ***Tranel, 2021***).

Here we investigate the evolutionary histories of mutations that have been previously demonstrated to confer target-site resistance in *A. tuberculatus*, focusing on TSR mutations within ALS and PPO. We infer the number of TSR mutational origins across populations and their distribution across the landscape, examining the signals left behind by both mutation and recombination. Specifically, we implement a method that infers the ancestral recombination graph (ARG) and that offers a powerful approach for inference of selective history, by providing near-complete information on relatedness among haplotypes (***Rasmussen et al., 2014***). Coupled with estimates of effective population size (N_e) through time based on coalescent rates across the genome (***Speidel et al., 2019***, ***2021***), these methods allow for powerful hypothesis testing on the role of standing variation versus new mutation. We assess heterogeneity in whether independent resistant lineages are associated with pronounced signals of selection based on a tree-based test and selective sweep signals, some of which may be mediated by intra- and inter-locus allelic interactions. We also examine signatures of the haplotype competition between common ALS resistance alleles, and the extent that extreme selection from herbicides on TSR mutations has impacted diversity across the genome. Our detailed population genomic analysis describing the repeatability of and heterogeneity in target-site herbicide resistance evolution advances our understanding of rapid adaptation of multicellular organisms to extreme selective pressure, while providing evolutionary informed priorities for agricultural weed management.

Types of target-site resistsance mutations

To test hypotheses about the origins of TSR in *Amaranthus tuberculatus*, we used whole-genome sequence information from 19 agricultural fields in the Midwestern US and Southwestern Ontario, Canada. It is important to note that these populations were obtained from fields where *A. tuberculatus* was only poorly controlled. Because our sampling is thus potentially over representing the frequency of resistance across the landscape, signatures of selection we observe across our samples are likely stronger than averages across these geographic regions would be.

Having previously characterized two types of target-site glyphosate resistance in these samples (coding sequence substitutions and gene amplification; ***Kreiner et al.*** (***2019***)), here we focus on all other characterized mutations known to confer resistance in the genus Amaranthus. We examined our sequence data for the presence of eight such substitutions in ALS, three in PPO, and one in photosystem II protein D1 (psbA). Across 152 individuals, we found six out of eight known ALS mutations, and one of the three known PPO mutations (Table 1). We did not find any mutation in psbA.

The nine unique PPO and ALS target-site resistant mutations occur at seven distinct codons, 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, with five or fewer instances, in contrast to the common Trp-574-Leu and Ser-653-Asn nonsynonymous substitutions in ALS, and the PPO ΔGly210 deletion (Table 1). Notably, the most common resistance mutation (referring to

Table 1. Number and frequency of resistant individuals and mutations at loci known to be causal for resistance to PPO and ALS herbicides, both totals, and within each agricultural region. Relative frequencies given in parentheses.

	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
Total # of inds. with TSR mutations	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)
Total # of TSR mutations	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)
TSR mutations in *Walpole*	0	12 (0.162)	12 (0.162)	1 (0.013)	0	0	0	0	0
TSR mutations in *Essex*	1 (0.013)	23 (0.288)	35 (0.438)	0	1 (0.013)	0	1(0.013)	2 (0.025)	3 (0.038)
TSR mutations in the *Midwest*	24 (0.16)	71 (0.47)	12 (0.08)	0	0	1 (0.007)	0	0	0

identity-by-state), Trp-574-Leu, is found in 53% of agricultural individuals, and the second most common, Ser-653-Asn, in 32% of individuals (Table 1). From these two most frequent ALS mutations alone, 74% of individuals sampled had acquired resistance to ALS-inhibiting herbicides. Accounting for rare ALS resistance mutations only increases this percentage to 75%, because these rare variants are almost exclusively found in individuals already harbouring one of the two common ALS mutations.

Regardless of the geographic region (within Essex County, Walpole Island, and the Midwestern US), multiple causal changes confer ALS resistance. Furthermore, nearly all populations (5/8 populations within the Midwestern US, 5/5 populations in Essex County, and 4/6 populations in Walpole) harbour multiple causal ALS mutations (Table 1). Thus, at just the level of these distinct mutational types, we observe genetic convergence in adaptation to ALS-inhibiting herbicides at global, regional, and population scales.

Regional selective sweep signals

To learn how and how often the individual mutations might have arisen, we first visualized regional selective sweep patterns at PPO and ALS genes—two genes that happen to be located 250 kb apart in the genome—with respect to the common Trp-574-Leu, Ser-653-Asn, ΔGly210 alleles. In particular, we assayed the extent to which selection from herbicides at these genes has led to reductions in diversity, and increases in homozygosity and linkage across the haplotype, as would be expected if TSR alleles have increased in frequency rapidly enough that recombination has yet to unlink these alleles from the background on which they arose. Corresponding selective sweep signals appear to be highly heterogeneous across geographic regions and across resistance mutations (Fig 1). The most pronounced selective sweep signal at the regional level is for the ALS Ser-653-Asn mutation, 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, Fig 1 top-left). The breadth of the impact of selection on local chromosome-wide linkage disequilibrium (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 these same individuals (***Kreiner et al., 2019***). This pattern of a hard selective sweep is also apparent in patterns of pairwise diversity and in LD of the focal Ser-653-Asn mutation with missense SNPs (r^2) (Fig 1, top middle/right). In contrast, the Trp-574-Leu mutation in Essex actually

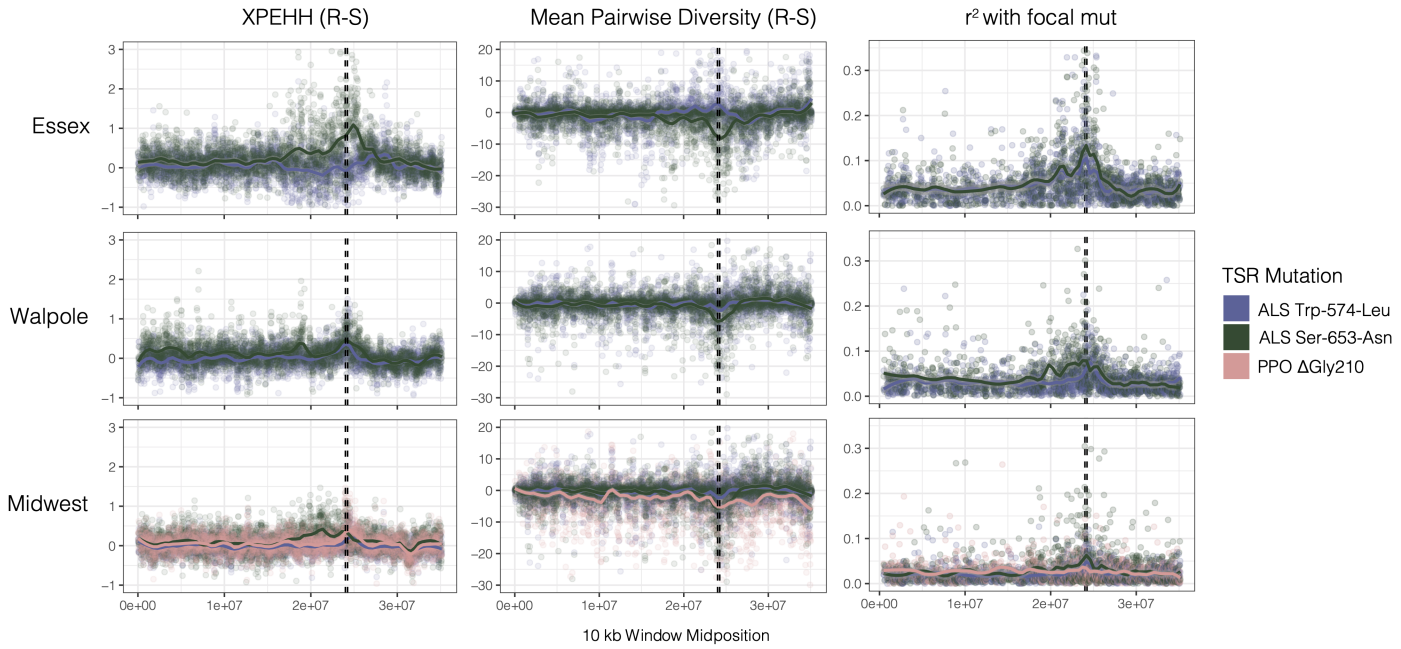

Figure 1. Sweep-scan summary statistics by geographic region. *Left*) Difference in integrated haplotype homozygosities (XPEHH) between haplotypes carrying the focal TSR mutation and susceptible haplotypes. *Middle*) Difference in mean pairwise diversity between haplotypes carrying the focal TSR mutation and susceptible haplotypes. *Right*) r2 of other missense mutations with focal TSR mutation based on genotype, rather than haplotype data. In all columns, dashed vertical lines denote PPO (left) and ALS (right) genes, which are only 250kb apart in the genome.

shows a slight excess of heterozygosity and excess diversity compared to susceptible haplotypes, but nearly as strong LD with other missense SNPs.

Selective sweep signals are much subtler in Walpole and especially in the Midwestern US compared to Essex, the latter likely influenced by broader population sampling (Table 1). In Walpole, ALS resistance shows a stronger signal of selection for Ser-653-Asn than for Trp-574-Leu, whereas signatures of selection for either mutation are almost completely lacking from the Midwestern US except for a slight peak in r^2 for ALS Ser-653-Asn. The PPO ΔGly210 mutation is found at considerable frequencies only in the Midwestern US, but regional sweep signals based on homozygosity, diversity, and LD are absent with respect to 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***; ***Patzoldt and Tranel, 2007***; ***Shoup et al., 2003***), and in congeners for the remaining mutations (***McNaughton et al., 2001***; ***Nakka et al., 2017***; ***Singh et al., 2018***; ***Whaley et al., 2004***)]. Thus, we set out to identify the extent to which repeated origins and gene flow have influenced regional signatures of selection, as well as identify key processes that may underlie heterogeneity in their recent evolutionary histories.

Inferring the genealogical history of target-site resistance mutations

We first took a gene tree approach to reconstruct the evolutionary history of TSR mutations, based on phased haplotypes inferred from performing the most up-to-date joint population and read-backed phasing methods (SHAPEIT4 (***Delaneau et al., 2019***); WhatsHap v1.0 (***Martin et al., 2016***)). We found that patterns of similarity among phased haplotypes at ALS and PPO (including 5 kb upstream and downstream of either gene) indicated numerous origins for every common resis-

tance mutation: PPO ΔGly210, ALS Trp-574-Leu, and Ser-653-Asn (Figure 2 - figure supplement 1). 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. Because recombination causes ancestral haplotypes to decay in size as they are passed down through time, linked sites may not necessarily have identical genealogies as a single mutational origin may be recombined onto distinct haplotypes. In cases such as this, ancestral recombination graphs (ARGs) can allow for more accurate inferences of genealogical history by generalizing the inference of coalescent history along a recombining unit (***Griffiths and Marjoram, 1996***, ***1997***; ***Hudson, 1983***).

We reconstructed the ARG for 20,000 SNPs encompassing both ALS and PPO genes (a 1 and 10 kb gene respectively, separated by 250 kb on the same chromosome) using ARGweaver (***Rasmussen et al., 2014***; ***Hubisz and Siepel, 2020***). We assessed the likelihood of the ARG inferences under varying constant recombination rates and over two time step parameters (Figure 2 - figure supplement 2 and 3). From our most likely parameter values (recombination = 10^{-8}, time steps = 30), and based on the MCMC sample that maximizes the likelihood of our data across 1,250 iterations, we extracted the tree corresponding to each focal TSR locus. For all three common TSR mutations, ALS Trp-574-Leu, ALS Ser-653-Asn, and PPO ΔGly210, we found evidence for multiple independent origins producing the same resistant variant—three for ALS Trp-574-Leu, two for ALS Ser-653-Asn, and two for PPO ΔGly210 (Figure 2A). Support for these origins was generally very high, with 5/7 origins showing that 100% of MCMC samples were consistent with each cluster of haplotypes being monophyletic. The two origins with less than full support were haplotypes harbouring the ALS Trp-574-Leu, where one high-frequency origin had 85% support ($^{105}/_{125}$ MCMC samples) and a low-frequency origin had 45% support ($^{56}/_{125}$), implying that occasionally haplotypes mapping to these origins belonged to other groupings across MCMC samples. The findings of multiple origins of identical resistance mutations, and most specific origin scenarios, are consistent even across less likely recombination rate parameterizations (Figure 2 - figure supplement 2 and 3). Compared to the gene trees based on the average mutational history across the PPO and ALS (Figure 2 - figure supplement 1 and 4), accounting for recombination has clearly further resolved the origins of these target-site resistance alleles. In comparison to susceptible haplotypes, resistant haplotypes at each origin are apparently more homozygous but retain some diversity, presumably driven by both mutation and recombination subsequent to their origin (Figure 2 - figure supplement 4, 5, 6).

When haplotypes belonging to distinct mutational lineages are mapped across populations (Fig 2B), it is clear that, despite the many independent origins, gene flow has also played a major role in the spread of resistance across the landscape. Haplotypes from the three most common origins of resistance to ALS herbicides—Trp-574-Leu #1, #2, and Ser-653-Asn #5 (corresponding to 39, 25, and 47 haplotypes, respectively)—were identified in 15, 10, and 14 populations. To test if mutational lineages were more geographically structured than expected given their frequency, we performed a permutation of haplotype assignment to a geographic region. All but two mutational lineages were consistent with expectations under this null, suggesting near panmixia between the Midwestern US and Ontario (Essex + Walpole) for most resistant lineages. The exceptions were ALS Trp-574-Leu #2, which is near exclusive to the Midwestern US (44/47 or 94% of samples; 95% CIs of regional permutations [0.383,0.638]), and Ser-653-Asn #5, which is at a much higher frequency in Ontario populations (87% found in either Essex or Walpole; 95% CIs of regional permutation [0.370, 0.630]), suggesting that these mutations arose locally in these geographic regions.

We next performed a tree-based test of non-neutral allele frequency change to examine whether TSR alleles have experienced consistent or shifting selection over their histories. Specifically, we implemented a tree-based statistic that relies on the order of coalescent events (***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). 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 are relative to all other susceptible lineages.

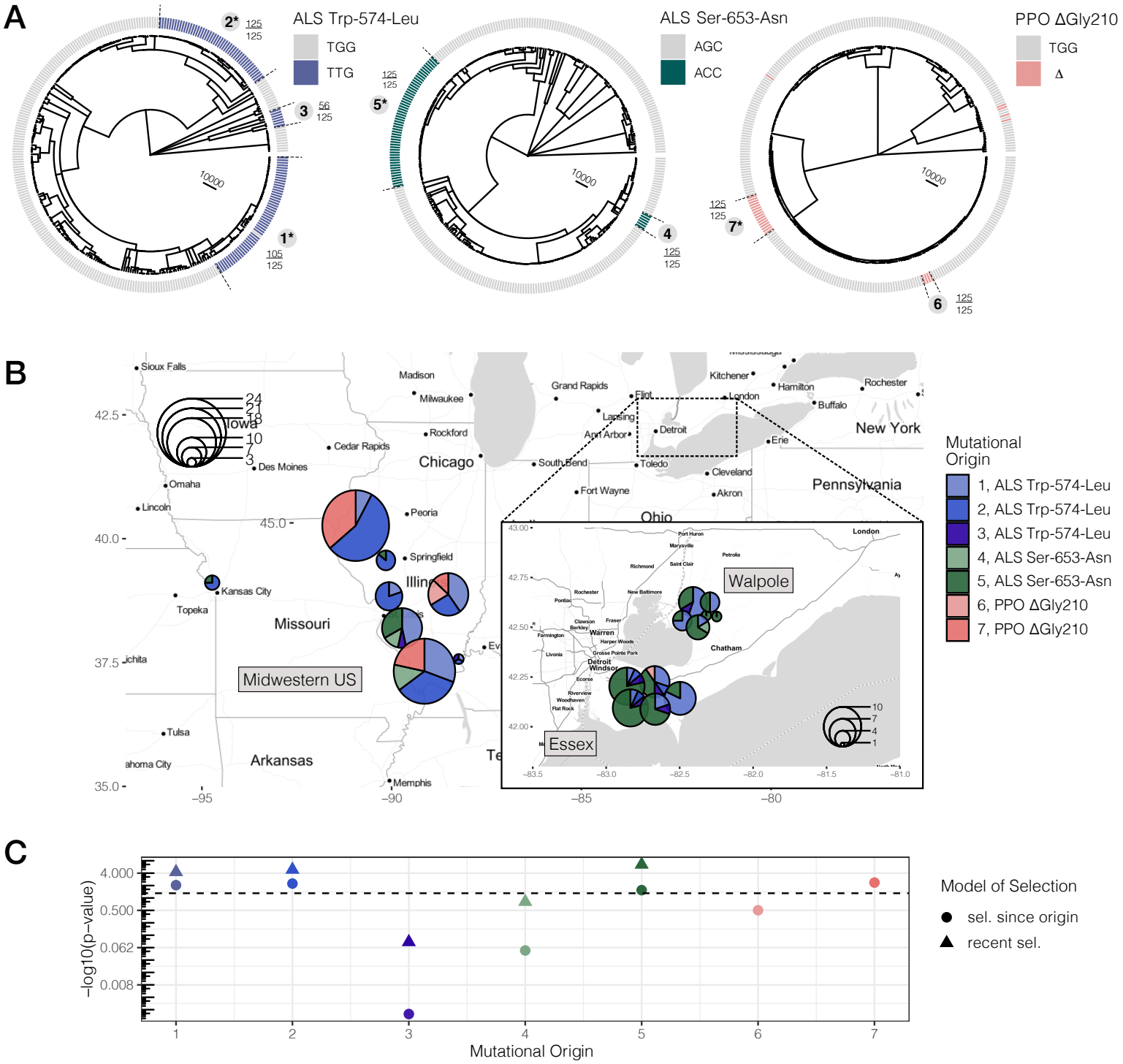

Figure 2. Repeated independent origins and range-wide distribution of three target-site resistance mutations, along with their associated significance of selection over two different timescales. **A**) Tree corresponding to ARGs estimated at focal TSR loci. Bold numbers around trees identify clusters of resistant haplotypes consistent with independent origins. 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, as in *C*. Support for monophyly for each origin across 125 samples of 1,250 MCMC iterations is depicted by the fraction found outside each cluster. **B**) Geographic distribution of haplotypes originating from distinct mutational lineages as inferred from trees extracted from ARGs at the focal TSR loci. TSR mutational lineages are found across numerous populations and agricultural regions, although regions show clear differences in the frequency of some mutations. **C**) Results of tree-based tests of non-neutral allele frequency change (***Speidel et al., 2019***) from each mutational origin of TSR under two alternative models of selection; selection on a mutation since its origin versus selection over more recent timescales (on the last 0.01% of the tree). The horizontal dashed line denotes the *p*-value cutoff of α=0.05.

Under the scenario of consistent selection since the origin of the mutation, four out of seven mutational origins we tested were significant at α=0.05 after a 5% false discovery rate (FDR) correction (Fig 2C, Figure 2 – source data 1). Since it arose, the Midwestern US high-frequency ALS Trp-574-Leu #2 variant showed the strongest signature of selection across all origins and all resistant loci ($p = 0.0058$), followed by the widespread high-frequency ALS Trp-574-Leu #1 variant ($p = 0.0089$), the common Midwestern US PPO variant #7 ($p = 0.00434$), and lastly, the high-frequency Ontario ALS Ser-653-Asn #5 variant ($p = 0.02831$).

In addition, a test for selection over the most recent 0.01% of the tree showed that for all three out of the four selected lineages that predated this time cutoff, there is even stronger evidence for selection on more recent timescales, after FDR correction (Fig 2C, Figure 2 – source data 1). The most obvious of these is the ALS Ser-653-Asn variant #5, which, while having the weakest evidence of consistent selection over its history, shows the strongest evidence of selection over recent timescales ($p = 3.44 \times 10^{-7}$). These tests therefore illustrate a strong role for fluctuating selection, intensifying over recent timescales.

While our previous test provides insight into the consistency of selection across the course of a mutational lineage's history, a conceptually related approach is to directly assess the role of resistance adaptation from standing genetic variation or new mutation based on allele age estimates relative to the onset of the selection. Allele age estimates depend greatly on the accuracy of effective population size estimates over the relevant evolutionary timescale. Namely, for herbicide resistance evolution, we posit that the relevant N_e is most likely the effective population size over the last half-century or less, corresponding with the introduction of agronomic pesticide regimes. While we have previously used $\delta a \delta i$ (***Gutenkunst et al., 2009***) to model species-wide demography in a two-epoch model and found a large population size expansion (historical epoch $N_e \approx 500,000$; recent epoch var. *rudis* $N_e \approx 5,000,000$; (***Kreiner et al., 2019***)), we now used Relate (***Speidel et al., 2019***, ***2021***) to infer effective population size through time from genome-wide tree sequence data on even more recent timescales. Historical N_e between 100 and 1,000,000 years ago appears to have stayed relatively consistent, with a harmonic mean of $\approx 63,000$ (standard error across chromosomes $\pm 7,000$ years) (Fig 3A). However, 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 (Fig 3A)—$80,000,000$ over the timescale of ALS herbicide use (approximately the last 40 years).

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 1,250 MCMC ancestral recombination graph iterations. We found considerable variation in the estimated age of resistance mutations across distinct lineages of the same mutation and across the three different TSR loci, according to haplotype groupings from the most likely ARG (Fig 3B). Across all mutational lineages, even for those with slightly lower support for monophyly (ALS Trp-574-Leu #1, #3), the 95% confidence interval of allele ages show these estimates are highly consistent across MCMC sampling of the ARG (Fig 3B). Our estimates of the age of the PPO ΔGly210 lineages are extremely recent and much less variable compared to ALS Trp-574-Leu and ALS Ser-653-Asn lineages (median PPO ΔGly210 age = 3.7 years [SE=0.2]; ALS Ser-653-Asn = 37.8 [SE=7.1]; ALS Trp-574-Leu = 37.2 [SE=106.3]). With ALS-inhibiting herbicides having been introduced in the early 1980s, only one lineage (ALS Trp-574-Leu #3) appears to substantially predate herbicide, with an estimated age of 350 years. However, this lineage is also the one with the least support for monophyly, which will upwardly bias estimates of allele age. Thus, while the exact timescales highly depend on accurate estimation of contemporary effective population sizes, the results are generally consistent with most mutations arising very recently after the onset of herbicide use.

In aggregate, our analyses have uncovered multiple independent origins of large-effect resistance mutations, along with heterogeneity in their evolutionary histories, from the timescale over which they have persisted to signatures of selection over varying timescales. The spread of these parallel origins across the landscape further allows us to observe how these alleles interact when

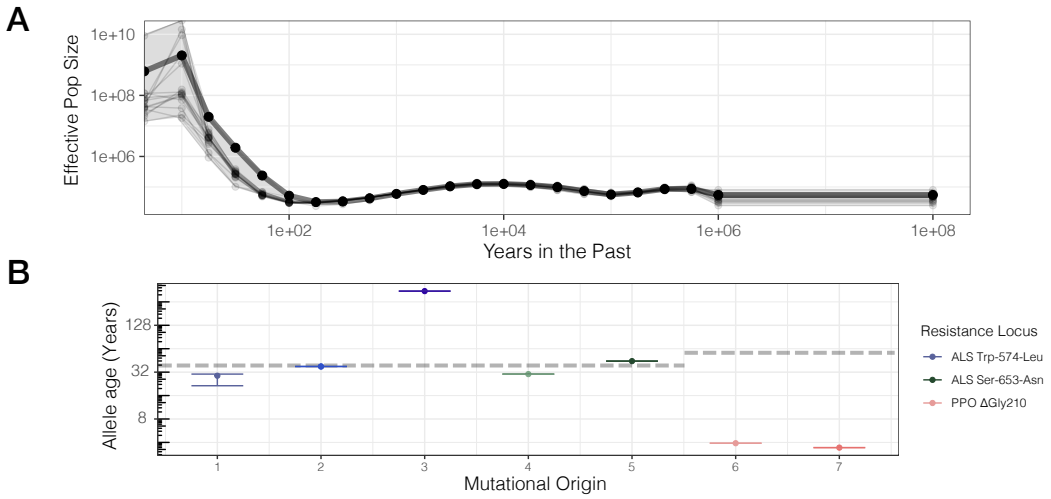

Figure 3. Contemporary population expansion of *A. tuberculatus* and corresponding ages of TSR variants. **A**) Relate-inferred effective population size through time, illustrating a remarkable population expansion occurring over the last 100 years. The bold line indicates results from genome-wide SNPs, 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[N_e] = 83,294,700). Horizontal dashed lines for ALS Trp-574-Leu and ALS Ser-653-Asn, and PPO origins represent the approximate onset of ALS and PPO herbicide use, respectively.

they meet (***Ralph and Coop, 2010***), and how this interaction may be modified by other alleles across the genome.

Haplotype competition and inter-locus interactions of TSR mutations

While 16 individuals harbour both the common ALS Trp-574-Leu and ALS Ser-653-Asn mutations, haplotype-level analyses indicate that no single haplotype harbours both mutations (Figure 2 - figure supplement 1, 4). This lack of double resistant haplotypes is a strong violation of expectations under linkage equilibrium ($\chi^2_{df=1} = 16.18$, $p = 5.77 \times 10^{-5}$), further suggesting that no or very little recombination has occurred between these sites. Given how globally common these resistance alleles are (Table 1; 53% and 32% of individuals harbour ALS Trp-574-Leu or Ser-653-Asn), their coexistence yet independence suggest that allelic competition may be important in adaptation to ALS-inhibiting herbicides.

Locally in Essex, where haplotypes carrying either the ALS Trp-574-Leu or ALS Ser-653-Asn mutations segregate at intermediate frequencies (29 and 44% respectively; Table 1), signed linkage disequilibrium (LD) between genotypes at these two sites is considerably negative ($r = -0.67$). In comparison to other non-synonymous SNPs at similar frequencies (minor allele frequency > 0.20) and separated by similar distances ($\leqslant 500$ bp) across the genome, this level of repulsion is unexpected (1-tailed $p = 0.033$; Figure 4 - figure supplement 1). However, this repulsion is not restricted to these two resistance mutations, which are only 237 bp apart. Rather, when we visualized the distribution of signed LD between all bi-allelic missense SNPs around the ALS locus, and the two TSR mutations, we observed a predominance of positive association with the focal ALS Trp-574-Leu mutation ($n = 19$), but a negative association with the ALS Ser-653-Asn mutation ($n = 34$), across an extended 10 Mb region (Fig 4A). This long-range repulsion is not seen for other pairs of loci that are outliers for particularly strong repulsion (Figure 4 – figure supplement 2). The divergence associated with these two TSR mutations segregating in Essex is also apparent with a genotype-based PCA—structure that is otherwise absent across the genome (Fig 4B, Figure 4 – figure supplement 3). Together, this indicates that selection for these alternative resistance variants from ALS-inhibiting

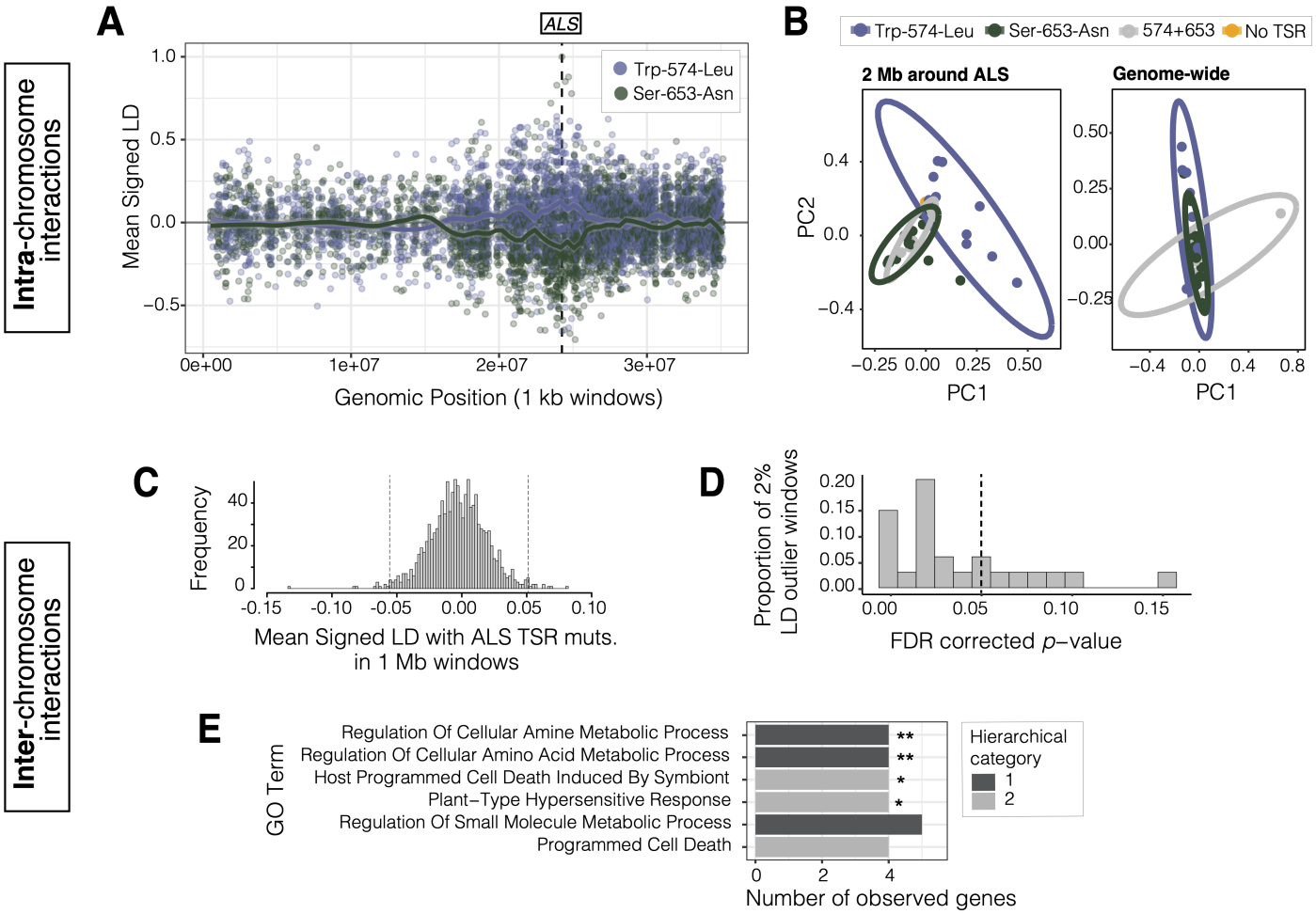

Figure 4. Signals of intra- and inter-chromosomal allelic interactions with target-site resistance mutations. **A**) The breadth of haplotype competition between TSR mutations, illustrated by repulsion linkage disequilibrium (opposite signed LD, *r*) between two target-site-resistance mutations and bi-allelic missense SNPs surrounding them on Scaffold 11 in Essex. Each point shows mean LD in non-overlapping 10 kb windows. A smoothing spline shows that missense SNPs tend to be in positive LD with ALS Trp-574-Leu but negative LD with ALS Ser-653-Asn in Essex. **B**) Signatures of population structure for 2 Mb around ALS compared to genome-wide, based on PCAs of genotypes in Essex. Ellipses represent 95% CIs assuming a multivariate distribution. **C**) Distribution of mean signed LD of ALS TSR resistance mutations (ALS Trp-574-Leu or Ser-653-Asn) with 1 Mb windows genome-wide in Essex, excluding the ALS containing Scaffold 11, with upper and lower 1% quantile indicated by dashed vertical lines. **D**) Distribution of p-values from top 2% of genome-wide windows with high absolute signed LD with ALS TSR mutations, from permuting individual assignment and recalculating LD 1,000 times. **E**) Significant enrichment of GO terms for the 590 genes mapping to the top 13, 1 Mb windows that show significantly extreme LD with ALS TSR mutations in Essex.

herbicides within Essex must occur through the competition of extended haplotypes up to ⅓ the size of this focal chromosome.

While the competition of haplotypes harbouring these TSR mutations may be important for shaping the distribution of resistance alleles across populations, the selective advantage of a given TSR haplotype may also depend on other modifier loci across the genome. In particular, we might expect that individuals that have withstood many generations of herbicide applications due to large-effect resistance mutations may have also accumulated compensatory and tolerance-conferring mutations across the genome (stacking of resistance alleles; (***Preston, 2003***; ***ffrench Constant et al., 2004***; ***Petit et al., 2010***; ***Busi et al., 2013***; ***Kreiner et al., 2021***). Considering that haplotype competition seems to have manifested itself in patterns of signed LD within the ALS containing chromosome, we posited that physically unlinked modifiers of resistance could be identified due to strong linkage with focal TSR loci.

To identify such interactions in populations in Essex, we calculated mean signed LD between focal ALS Trp-574-Leu or ALS Ser-653-Asn mutations and bi-allelic missense alleles in 1 Mb non-overlapping windows across the genome in Essex (Fig 4C). We then focused on the upper 1% and lower 1% of windows with particularly extreme signed LD with either TSR mutation (24/1,156 1-Mb windows). On each outlier window, we performed a permutation in which we randomized TSR allele assignment among individuals 1,000 times, to test whether observed LD with resistance genotypes was more extreme than expected. 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 ALS Trp-574-Leu (1-tailed $p < 0.0001, r = +0.068$) but negative signed LD with respect to the ALS Ser-653-Asn mutation ($r = -0.132$), consistent with repulsion between TSR alleles. Upon further inspection, this 1 Mb region is directly centred 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 24 outlier windows, 13 had p-values consistent with significantly extreme LD with TSR loci after FDR correction at $\alpha = 0.05$ (Fig 4D). These 13 windows included 348 genes, 120 of which have *Arabidopsis thaliana* orthologs. These 120 orthologs 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 (Fig 4E), seemingly directly related to the function of ALS—amino acid synthesis. Two particularly interesting examples from our set of genes in strong inter-chromosome LD with ALS resistance mutations are those encoding *GCN2* (general control non-repressible 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 (***Baena-González et al., 2007***; ***Chen et al., 2017***; ***Fujiki et al., 2001***).

Discussion

The application of herbicides to manage agronomically important weeds has led to one of the best-studied examples of parallel evolution outside the laboratory, with target-site-resistance mutations to ALS-inhibiting herbicides identified in more than 150 species (***Heap, 2021***). We have studied the evolution of resistance 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 and Canada, *A. tuberculatus*. We found rampant evidence for both independent origins and gene flow, competition among resistant haplotypes, and the interaction of large-effect TSR mutations with physically unlinked alleles with resistance-related functions. These results paint a picture of the rise, spread, and fate of adaptive alleles in the face of extreme selection, with important implications for the management of herbicide-resistant agricultural weeds.

Repeated origins and the spread of resistance via gene flow

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 in the sampled population were found at seven distinct codons, with nine distinct variants, three of which are common. These mutations have arisen repeatedly seven times across our sampled populations, based on ARG inference (Table 1, Figure 2), consistent with the largely soft selective sweep signals we observe at a regional scale (Figure 1). ARGs have seen limited implementation outside of human populations 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 comparison to gene trees that illustrate the average coalescent history of

these genomic regions (Fig 2 - figure supplement 1), we show that ARGs are extremely valuable for further resolving independent origins of adaptive mutations.

From a mutation-limited view of adaptation, the extent of parallelism in target-site herbicide resistance that we observe here is particularly extreme. However, given the estimate of North American *A. tuberculatus* $4N_e\mu \approx \Theta_\pi$ for putatively neutral sites (Kreiner et al., 2019), a new TSR mutation at any of the eight adaptive ALS mutations should arise at a rate of $2N_e\mu$—once every six generations (i.e. $^{0.041}/_{2} \times 8$ known TSR loci $= 0.164$ TSR mutations per generation; see also (***Charlesworth, 2009***; ***Karasov et al., 2010***). Furthermore, the estimated Θ_π we use to infer the rate of new mutations may even be an underestimate given contemporary population size, which may be closer to census size than long-term estimates of Ne from neutral polymorphism, should determine the mutational supply for rapid adaptation under models of evolutionary rescue (***Bell, 2013***; ***Karasov et al., 2010***; ***Kreiner et al., 2018***; ***Neve et al., 2014***). Indeed, if we modify this value to reflect the contemporary estimate of Ne over the last 50 years ($\approx 8 \times 10^7$) and assume an *A. thaliana* mutation rate of 7×10^{-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. From these inferences, parallelism in simple target-site herbicide resistance adaptation in *A. tuberculatus* appears to be on par with prokaryotic adaptation and pesticide resistance adaptation in *Drosophila melanogaster*, where population sizes on the order of $\Theta \approx 1$ facilitate 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 (Fig 2B). The widespread impact of gene flow is further consistent with our inference of near panmixia for all but two resistant lineages, although our permutation test is limited in power for rare origins. 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 (Fig 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 (***Sauer, 1957***), 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.

While the role of repeated origins and widespread gene flow we characterized here fit well with the cosmopolitan and highly convergent nature of herbicide resistance adaptation, the patterns we observe may be influenced by the sensitivity of ARG inference to both phasing quality and fine-scale recombination rate variation. We performed a two-step phasing method, performing population-level phasing with SHAPEIT4, a powerful up-to-date method (***Delaneau et al., 2019***), after performing read-backed phasing with WhatsHap (***Martin et al., 2016***). Nonetheless, phase switching remains a challenge for haplotype inference in naturally occurring populations in lieu of long-read population resequencing. Phase-switching between haplotypes is likely to be interpreted as a recombination event during ARG inference, however, by explicitly modelling how these "recombinational" units relate to one another, ARG inference should still be better able to resolve independent origins of adaptive haplotypes than traditional reconstruction methods. To adjust for the potential phase-switching that may artificially inflate recombination rates, we ran ARGweaver over three magnitudes of recombination rate values ($r = 10^{-7}$ to 10^{-9}), as well as two resolutions of discrete time steps ($t = 20, 30$) (Figure 2 - figure supplement 2), to find the parameter values that maximized the likelihood of our phased data ($r = 10^{-8}$ and $t = 30$). Even across less likely parameter values, we find consistent support for multiple origins across all TSR loci (Figure 2 - figure supplement 3).

The selective history of target-site resistance lineages

Detrimental effects of monogenic 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 were able to evaluate evidence for TSR alleles that predated the onset of herbicide usage, and thus learn about how long these alleles persist, by rescaling allelic age estimates by the geometric mean effective population size estimate over the last 50 years (Fig 3). We found evidence for one mutational lineage vastly predating the onset of ALS herbicide use, arising nearly 350 years ago; however, this lineage only showed 45% support for a monophyletic origin across MCMC iterations of the ARG inference. By and large, our results suggest that the ALS resistance mutations we have sampled arose repeatedly, soon after the introduction of ALS-inhibiting herbicides in the 1980s. The haplotypes on which they arose have since been subject to widespread gene flow, which can facilitate a rapid response to selection across a large geographic range. In comparison to these non-synonymous substitutions within the ALS gene, the single-codon deletion that confers resistance to PPO-inhibiting herbicides appears to be much younger, estimated to have only arisen within the last 3-4 years prior to population sampling.

While these allele age estimates provide a powerful test of the extent of adaptation from standing variation versus from new mutations, and the timescale over which resistance mutations may persist, these estimates are an approximation based on the geometric mean Ne over the last 50 years, and do not fully account for the monumental population expansion this species shows. Furthermore, this rescaling depends on the accuracy of our N_e estimate through time, as inferred by Relate (***Speidel et al., 2019***). Previously, we have also implemented δaδi (***Gutenkunst et al., 2009***), which uses the site frequency spectrum, to infer demographic history in these samples and similarly found evidence for large recent population expansion in a two epoch model (***Kreiner et al., 2019***). Relate is a better method for N_e inference in this scenario, however, as it provides estimates of N_e on contemporary timescales (0-100 years ago). Further, it has been shown to be accurate in the face of phasing error, with only a slight overestimation of Ne on recent timescales (***Speidel et al., 2019***). Nonetheless, this slight bias towards larger N_e estimates in the face of phasing error may suggest that our rescaled allele ages tend to underestimate the true age of these mutations.

Given the challenges of allele age estimation, we also used a conceptually related approach, testing for evidence of selection over two different timescales with a tree-based statistic that is robust to population size misspecification. Reassuringly, this tree-based test clearly shows evidence of selection since a mutation arose (consistent with *de novo* adaptation) for the history of four out of seven lineages. However, this test further demonstrates that these lineages are even more likely to have been under selection on more recent timescales, with rank order shifts in support across resistant lineages as compared to support for consistent selection over their history. Thus, while by our estimates, adaptation to PPO- and ALS-inhibiting herbicides relies predominantly on *de novo* mutation, spatially and temporally varying selection has resulted in muted signals of selection over the course of many mutational lineages' histories—selection that has intensified over timescales even more recent than the onset of herbicide use in some geographic regions.

While the intensification of herbicide use over the last half-century may be one source of temporally varying selection, rotating herbicide and cropping regimes may also contribute to fine-scale fluctuations in selection for or against particular TSR mutations. For example, in corn and soy production systems (where the focal crop alternates each season), 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, the lower net-positive selection from PPO-inhibiting herbicides along with their recent resurgence in popularity (***Tranel, 2021***) may explain the more recent origins and thus shorter persistence of the PPO ΔGly210 alleles. Similarly, while the ALS Trp-574-Leu mutation tends to confer resistance broadly across all ALS herbicides, the ALS Ser-653-Asn mutation confers resistance to fewer types of ALS-inhibiting herbicides, which also happen to be used more commonly in soy (***Pat-***

zoldt and Tranel, 2007). This may contribute to the relatively lower frequency of ALS Ser-653-Asn compared to ALS Trp-574-Leu across the *A. tuberculatus* range, or even suggest that a lack of both focal crop and herbicide rotation facilitated strong recent selection on the high-frequency Ontario ALS Ser-653-Asn lineage.

Intra and inter-locus allelic interactions shape the history of TSR mutations

The outcome of parallel adaptation in a continuous species range has been thoroughly described by *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, beneficial haplotypes will compete on their way to fixation (*Ralph and Coop, 2010*). Given our evidence for highly parallel TSR adaptation across the range, we expect that this scenario fits the evolution of resistance particularly well. The widespread gene flow we observe sets up a scenario where independent origins of TSR mutations have now met and must interact. Under such a scenario, further background-dependent fitness effects, additive and epistatic interactions with other loci, both physically linked and unlinked ones, may shape the success of particular mutational lineages.

The coexistence of mutations of independent origin under strong selection in a given locale, notably of the ALS Trp-574-Leu and ALS Ser-653-Asn mutations in Essex, has allowed us to observe the signature of such intra-locus interactions. The 10 Mb stretch of repulsion and strong haplotypic divergence we observe (Fig 4A, B) likely reflects the unique spatial origins of each focal ALS TSR mutation and local selection. However, these mutations now co-occur across many agricultural regions. The ongoing repulsion of these alleles, combined with selection being constrained to act on independent allelic combinations (*Otto, 2021*), means that further adaptation to ALS-inhibiting herbicides must occur through the competition (sensu *Mather* (*1969*)) of these extended resistance haplotypes nearly 10 Mb long.

One question that arises based on our observations of haplotypic competition is why recombination has not eroded the signal of LD, and further, why we do not observe a single haplotype with both ALS mutations? The former is clearly not because of a lack of opportunity for recombination, since trans-heterozygous individuals that carry both mutations in opposite phases are not uncommon, and both mutations segregate at considerable frequencies. However, it is possible that recombinant double mutants are only rarely generated through recombination. The local LD-based population recombination rate estimate of $\rho = 4N_e R = 0.0575$ in a region of 100 kb on either side of ALS implies 3.4 new recombination events per generation in the distance between these two mutants ($237\text{bp} \times \frac{0.057}{4}$). Accounting for the local frequencies of ALS Trp-574-Leu and ALS Ser-653-Asn alleles (0.29×0.44), in a panmictic population, ≈ 1 of these recombination events should generate a double resistant mutant every other generation. However, recombination is preferred in promoter regions (*Kent et al., 2017*; *Sandler et al., 2020*), suggesting that these calculations may overestimate intragenic levels of recombination and thus the opportunity to generate recombinant resistant haplotypes.

Understanding the fitness consequences of these alleles, both separately and in combination, will determine how this competition will be resolved. If one allele confers significantly greater resistance (and assuming similar costs in the absence of herbicides), we may expect it to eventually reach fixation, outcompeting the alternate haplotype. Alternatively, if the alleles are selectively neutral with regard to each other, stochastic processes may predominate (*Ralph and Coop, 2010*). Given enough time, however, recombination should create haplotypes carrying both TSR mutations. It is possible that the joint effect of non-synonymous mutations at codon 574 and 653 on ALS protein structure results in negative epistasis, which would keep such double resistant haplotypes rare (as observed for antibiotic resistance alleles (*Porse et al., 2020*). Alternatively, additive or positive epistatic effects between these mutations would favour fixation of a double mutant haplotype, suggesting that the current observed haplotypic competition is in fact a form of Hill-Robertson in-

terference (where the rate of adaptation is slowed due to linkage) which can eventually be resolved through recombination (***Cooper, 2007***; ***Hill and Robertson, 1966***; ***Otto, 2021***).

While we find that intra-chromosomal interactions such as competition between alleles have influenced the selective trajectory of individual TSR alleles, we were also interested in the extent to which interactions with physically unlinked loci may have facilitated herbicide resistance evolution. We find evidence that selection on Essex haplotypes containing ALS TSR mutations has likely been affected by such second-site modifiers (Fig 4C-E). We find particularly extreme signed LD between TSR mutations and alleles on different chromosomes. LD between resistance alleles on different chromosomes has been interpreted as epistasis (***Gupta et al., 2021***)), but LD between alleles that are not physically linked may also result from the effects of additive alleles with correlated responses to selection (e.g. the stacking of resistance alleles; ***Busi et al.*** (***2013***)).

Alleles in windows on different chromosomes with the strongest evidence of linkage with ALS target-site resistance mutations function in biological processes related to known organismal responses to ALS herbicides. 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 (***Orcaray et al., 2011***; ***Trenkamp et al., 2009***; ***Zhao et al., 2018***; ***Zulet et al., 2013***). We observe significant enrichment for functions in both amino acid production and cell death. These alleles may thus provide additional levels of tolerance on the large-effect TSR background, by compensating for homeostatic disturbances caused by ALS exposure or the potential costs of large-effect resistance mutations (as seen for antibiotic resistance (***Craig MacLean et al., 2010***)).

Conclusion

In conclusion, adaptation to herbicides and the emergence of well-characterized target-site resistance mutations provide a powerful system for characterizing rapid and repeated evolution in plant populations, as well as the consequences of extreme selection on local and genome-wide patterns of diversity. Studies of herbicide 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., 2021***); reviewed in (***Délye, 2013***; ***Powles and Yu, 2010***), costs of adaptation under fluctuating environments (***Vila-Aiub, 2019***; ***Vila-Aiub et al., 2009***), and mutational repeatability (e.g. ***Menchari et al.*** (***2006***); ***Heap*** (***2014***)) (see ***Baucom*** (***2019***)). The work here contributes to this literature by characterizing not only the parallel origins and spatial distribution of target-site-resistance alleles across a broad collection of agricultural populations, but also heterogeneity in their evolutionary history and key contributing processes, such as fluctuating selection, haplotype competition, and cross chromosomal linkage with putative resistance alleles.

From a practical perspective, this work highlights *A. tuberculatus* as a worst-case scenario for controlling the evolution of herbicide resistance and containing its spread. Large census and effective population sizes facilitate extreme convergence for repeated selection of identical nucleotide polymorphisms conferring resistance. One key priority for thwarting new resistance mutations from arising and spreading should thus be containing *A. tuberculatus* population sizes as much as possible. Our findings suggest that most resistance mutations are of very recent origin and that they can persist for several decades with average herbicide usage. Across growing seasons, our best hope is to reduce the fitness advantage of resistant types by decreasing reliance on herbicides, rotating herbicide chemical compounds and mode-of-actions, and using herbicide mixtures. However, one must also be wary of neighbouring weed populations. Gene flow of resistance mutations across the range is widespread, facilitating a rapid response to selection from herbicides, potentially even in naive populations. In particular, we find strong regional patterns in the distribution of resistant lineages, suggesting coordinated herbicide management regimes across farms,

land-use planning, and hygienic machine-sharing will be important tools for efficient control of herbicide-resistant weeds.

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 Midwestern US (collected 2010), and from newly infested counties in Ontario, Canada, Walpole Island and Essex County (collected 2016). The eight Midwestern US populations previously had been surveyed for resistance to glyphosate (***Chatham et al., 2015***). Ten 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 convergent 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., 2021***).

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 for 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) (***Yu and Powles, 2014***), 3 PPO amino acids (***Giacomini et al., 2017***; ***Rousonelos et al., 2012***; ***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***). To assay these mutations in our samples, we referred to the literature on previously verified TSR mutations, extracting the sequence surrounding a given focal TSR mutation, and BLAST (***Altschul et al., 1990***) searched our reference genome to locate its position.

To phase genotypes into haplotypes, we first used WhatsHap (***Martin et al., 2016***), which performs readback phasing, and passed these phase sets to SHAPEIT4 (***Delaneau et al., 2013***) which further phases haplotypes based on population-level information. 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. SHAPEIT4 also requires a genetic map; with no 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 ($\frac{100}{4N_eR}$; $N_e = 500,000$), and used a monotonic spline to extrapolate genetic distance to each SNP in our VCF. We provided SHAPEIT4 an effective population size estimate of 500,000, inferred from previous demographic modelling in $\delta a \delta i$ (***Kreiner et al., 2019***).

Tree inference

Gene trees were inferred based on haplotypes within focal target-site genes (ALS and PPO), and 5 kb on either side around them. This resulted in 296 SNPs and 622 SNPs for analysis for the ALS and PPO trees respectively. Using the phased data around these genes, we first converted each phased haplotype to FASTA format, performed realignment with clustal omega (***Madeira et al., 2019***), and then ran clustal-w2 (***Thompson et al., 1994***) to infer the maximum likelihood tree, once for each

gene, with 1,000 bootstraps. We then plotted mutational status for each focal TSR mutation (ALS Trp-574-Leu, ALS Ser-653-Asn, and PPO ΔGly210) for each tip of both gene trees using ggtree (***Yu, 2020***).

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 across our phased haplotypes. We used the settings -N (effective population size) 500,000 -m (mutation rate) 1.8e-8 –ntimes (estimated timepoints) 20 –maxtime (max coalescent time) 100e3 -c (bp compression rate) 1 -n (MCMC samples) 1250. We used an effective population size of 500,000, based on the best fitting demographic model previously inferred from this dataset with $\delta a \delta i$ (***Gutenkunst et al., 2009***; ***Kreiner et al., 2019***). To account for any bias introduced by phasing error in the form of inflated recombination rates, we ran multiple iterations of ARGweaver at varying constant recombination rates, from $r = 7 \times 10^{-7}$ to 7×10^{-9}, drawing our inferences from the parameter values that maximized the likelihood of observing our data ($r = 7 \times 10^{-8}$). Similarly, we also tested out two parameter values of –ntimes (20, 30), and used t=30 for our inference, together with $r = 7 \times 10^{-8}$. We then extracted the most likely ARG sample from the MCMC chain (sample 1240), 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 all 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 each origin. ARGweaver runs on phased bi-allelic SNP data. Therefore to obtain a tree for the PPO ΔGly210 deletion, we recoded the VCF entry at this indel as a SNP, changing the reference allele at this site to a single locus record matching the reference call at the start position (G) and an alternate single base-pair call (A).

As a test for panmixia, we performed a permutation test of the proportion of a mutational lineage (i.e. haplotypes belonging to a particular origin) mapping to Ontario (Essex and Walpole) compared to the Midwest. In particular, we randomized haplotype assignment to either Ontario or the Midwest, and recalculated the proportion of haplotypes belonging to each origin found in each region, 1000x. We then calculated the 95% CI of the proportion of haplotypes mapping to either Essex or the Midwest, as the null expectation under panmixia. When our observed value for a given origin exceeded the null expectation, we took this as significant evidence for stratification in the geographic distribution of a mutational lineage.

Coalescent tree-based tests for selection

RELATE (***Speidel et al., 2019***, ***2021***), a scalable method for estimating ARGs 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). We did this on an origin-by-origin basis, comparing the rate of allele frequency change across the tree for a focal resistant lineage compared to all other susceptible haplotypes across the tree. Since this statistic is simply based on the order of coalescents, rather than branch lengths, it should be robust to misspecified effective population sizes used to infer our ARG ***Speidel et al.*** (***2019***). Since RELATE assumes an infinite sites model and thus is unsuitable 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 prob-

ability 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-side p-value is sufficiently small (i.e. $\alpha < 0.05$), 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

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), re-estimated branch lengths with this updated mutation rate (ReEstimateBranchLengths), and then lastly estimated population size through time (EstimatePopulationSize.sh). Population sizes were estimated from 0 years ago to 10,000,000 years ago, in epochs timesteps of $10^{0.25}$ years, to obtain particularly fine-scale estimates in the recent past.

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 et al., 2007***), in this case, the difference in the integrated extended haplotype homozygosity between resistant and susceptible haplotypes, and mean pairwise difference estimates. For both of these statistics, we provided LD-based recombination maps, inferred from LDhat, as described in ***Kreiner et al.*** (***2019***). Because some individuals in Essex always carried at least one resistant ALS haplotype through either mutations at ALS Trp-574-Leu or ALS Ser-653-Asn, to compare patterns of selection associated with resistance and susceptibility, these statistics were calculated at the haplotype, rather than individual level. For each independent origin as inferred from ARGweaver, we similarly inferred XPEHH, as well as H12 (***Garud and Rosenberg, 2015***,?) across the chromosome containing ALS.

We used plink (v1.90b3.46) (***Purcell et al., 2007***) to calculate both signed LD (*r*) between each focal TSR mutation and missense mutations on the same chromosome, and with all other bi-allelic missense SNPs across the genome. We used nonsynonymous 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 nonsynonymous SNPs in Figure 4 - figure supplement 4. We performed these calculations with respect to a given TSR mutation by using the –ld-snp options to specify a focal mutation. To visualize patterns of signed LD between TSR mutations and other missense SNPs, we split the genome into non-overlapping 10 kb windows and calculated the average LD among all SNPs in each window. All LD calculations were polarized by rarity (e.g. minor alleles segregating on the same haplotypes were regarded as being in positive LD). In Essex, despite being considerably common, both ALS 574 and ALS 653 had a frequency less than 50%, so LD values between all missense alleles, and both these focal TSR mutations are directly comparable. As another visualization of the haplotype structure in this region, we conducted two genotype-based PCAs using the R package SNPrelate (***Zheng and Wiehe, 2019***), one for genotypes spanning 2 Mb around ALS, and one on genome-wide genotypes.

To test whether the negative LD we observed between ALS Trp-574-Leu and ALS Ser-653-Asn was particularly extreme in Essex, we compared this value to pairs of either nonsynonymous or synonymous SNPs of similar frequency (minor allele frequency > 0.20) and physical distance apart (<500bp). To test whether the top 2% of 1 Mb windows of nonsynonymous SNPs with particularly low or high signed LD with ALS TSR mutations (either ALS Ser-653-Asn or ALS Trp-574-Leu) was significantly different from the null expectation, we used a permutation approach whereby we randomly shuffled the assignment of the focal ALS TSR mutation between all individuals and calculated mean LD (with respect to the permuted TSR mutations) in the window of interest. We repeated this permutation 1,000 times to generate a null distribution for comparison to the real average signed LD value of each region. This permutation test explicitly evaluates whether a TSR mutation and missense mutations in a focal window are more likely to be found together in the specific set of individuals containing the focal TSR mutation than any other set of individuals of the same size. Thus, this test is robust to variance in data quality across windows. The proportion of permuted observations with a mean absolute signed LD exceeding the observed signed LD was taken as the two-tailed p-value for cross-chromosome LD. Lastly, we found the intersect of these windows with the closest gene according to our genome annotation, and found their *A. thaliana* orthologs (***Emms and Kelly, 2015***). We used the set of *A. thaliana* orthologs found across all 13 significantly enriched 1 Mb windows in a Gene Ontology (GO) Enrichment analysis for biological processes.

Acknowledgements

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

References

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

Baena-González E, Rolland F, Thevelein JM, Sheen J. A central integrator of transcription networks in plant stress and energy signalling. Nature. 2007 Aug; 448(7156):938–942.

Baker HG. The Evolution of Weeds. Annu Rev Ecol Syst. 1974; 5:1–24.

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

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, 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.

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.

Cooper TF. Recombination speeds adaptation by reducing competition between beneficial mutations in populations of Escherichia coli. PLoS Biol. 2007 Sep; 5(9):e225.

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.

Delaneau O, Zagury JF, Robinson MR, Marchini JL, Dermitzakis ET. Accurate, scalable and integrative haplotype estimation. Nat Commun. 2019 Nov; 10(1):5436.

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.

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.

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.

Garud NR, Rosenberg NA. Enhancing the mathematical properties of new haplotype homozygosity statistics for the detection of selective sweeps. Theor Popul Biol. 2015 Jun; 102:94–101.

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.

Green JM. Review of Glyphosate and ALS-Inhibiting Herbicide Crop Resistance and Resistant Weed Management. Weed Technol. 2007 Jun; 21(2):547–558.

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, The International Herbicide-Resistant Weed Database; 2021.

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

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.

Kent TV, Uzunović J, Wright SI. Coevolution between transposable elements and recombination. Philos Trans R Soc Lond B Biol Sci. 2017 Dec; 372(1736).

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 genome-wide signatures of glyphosate resistance in Amaranthus tuberculatus. Mol Ecol. 2021; .

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.

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.

Liu J, Davis AS, Tranel PJ. Pollen Biology and Dispersal Dynamics in Waterhemp (Amaranthus tuberculatus). Weed Sci. 2012; 60(3):416–422.

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.

Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 2019 Jul; 47(W1):W636–W641.

Martin M, Patterson M, Garg S, Fischer SO, Pisanti N, Klau GW, Schöenhuth A, Marschall T. WhatsHap: fast and accurate read-based phasing; 2016.

Mather K. Selection through competition. Heredity. 1969 Nov; 24(4):529–540.

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.

Mazur BJ, Falco SC. The development of herbicide resistant crops. Annu Rev Plant Physiol Plant Mol Biol. 1989 Jun; 40(1):441–470.

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, 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.

Otto SP. Selective Interference and the Evolution of Sex. J Hered. 2021 Mar; 112(1):9–18.

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.

Porse A, Jahn LJ, Ellabaan MMH, Sommer MOA. Dominant resistance and negative epistasis can limit the co-selection of de novo resistance mutations and antibiotic resistance genes. Nat Commun. 2020 Mar; 11(1):1199.

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.

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.

Sandler G, Wright SI, Agrawal AF. Using Patterns of Signed Linkage Disequilibria to Test for Epistasis in Flies and Plants; 2020.

Sauer J. Recent Migration and Evolution of the Dioecious Amaranths. Evolution. 1957; 11(1):11–31.

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.

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.

Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994 Nov; 22(22):4673–4680.

Tranel PJ. Herbicide resistance in Amaranthus tuberculatus†. Pest Manag Sci. 2021 Jan; 77(1):43–54.

Tranel PJ, Trucco F. 21st-century weed science: a call for Amaranthus genomics. Weedy and invasive plant genomics. 2009; p. 53–81.

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

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

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

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

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

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

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

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

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

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

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

Yu G. Using ggtree to Visualize Data on Tree-Like Structures. Curr Protoc Bioinformatics. 2020 Mar; 69(1):e96.

Yu Q, Powles SB. Resistance to AHAS inhibitor herbicides: current understanding. Pest Manag Sci. 2014 Sep; 70(9):1340–1350.

Zhao LX, Jiang MJ, Hu JJ, Zou YL, Gao S, Fu Y, Ye F. Herbicidal activity and molecular docking study of novel PPO inhibitors. Weed Sci. 2020 Nov; 68(6):565–574.

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 herbicide tolerance via GCN2-independent regulation of amino acid homeostasis. Autophagy. 2018 Jan; 14(4):702–714.

Zheng Y, Wiehe T. Adaptation in structured populations and fuzzy boundaries between hard and soft sweeps. PLoS Comput Biol. 2019 Nov; 15(11):e1007426.

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

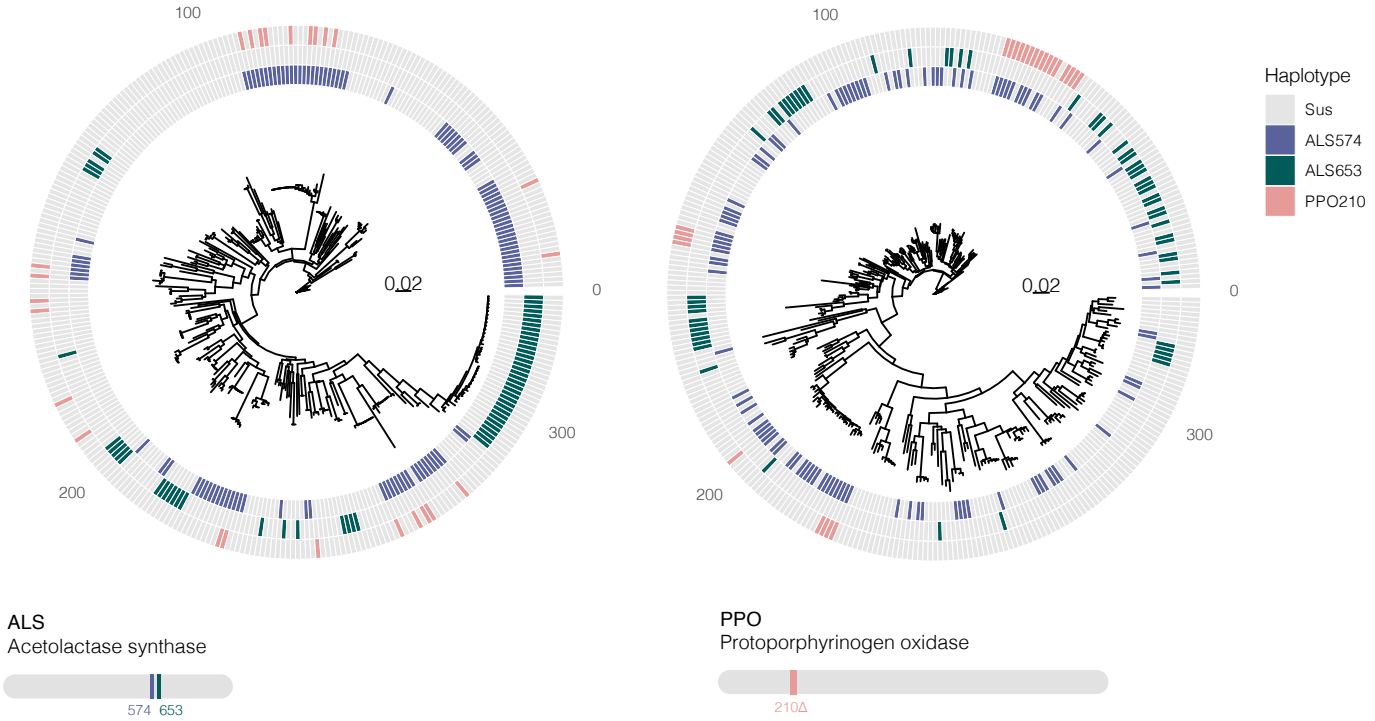

Figure 2 - figure supplement 1. Bootstrapped gene trees of ALS (3 kb) and PPO (10 kb) (CDS +1 kb on either side) alongside TSR mutations across all 162 individuals. Coloured grids indicate a haplotype harbouring a focal resistance SNP, with the SNP legend inserted within the grid (ALS Trp-574-Leu mutations in green, ALS Ser-653-Asn mutations in purple, PPO ΔGly210 deletions in pink). Multiple origins are apparent for all resistance mutations.

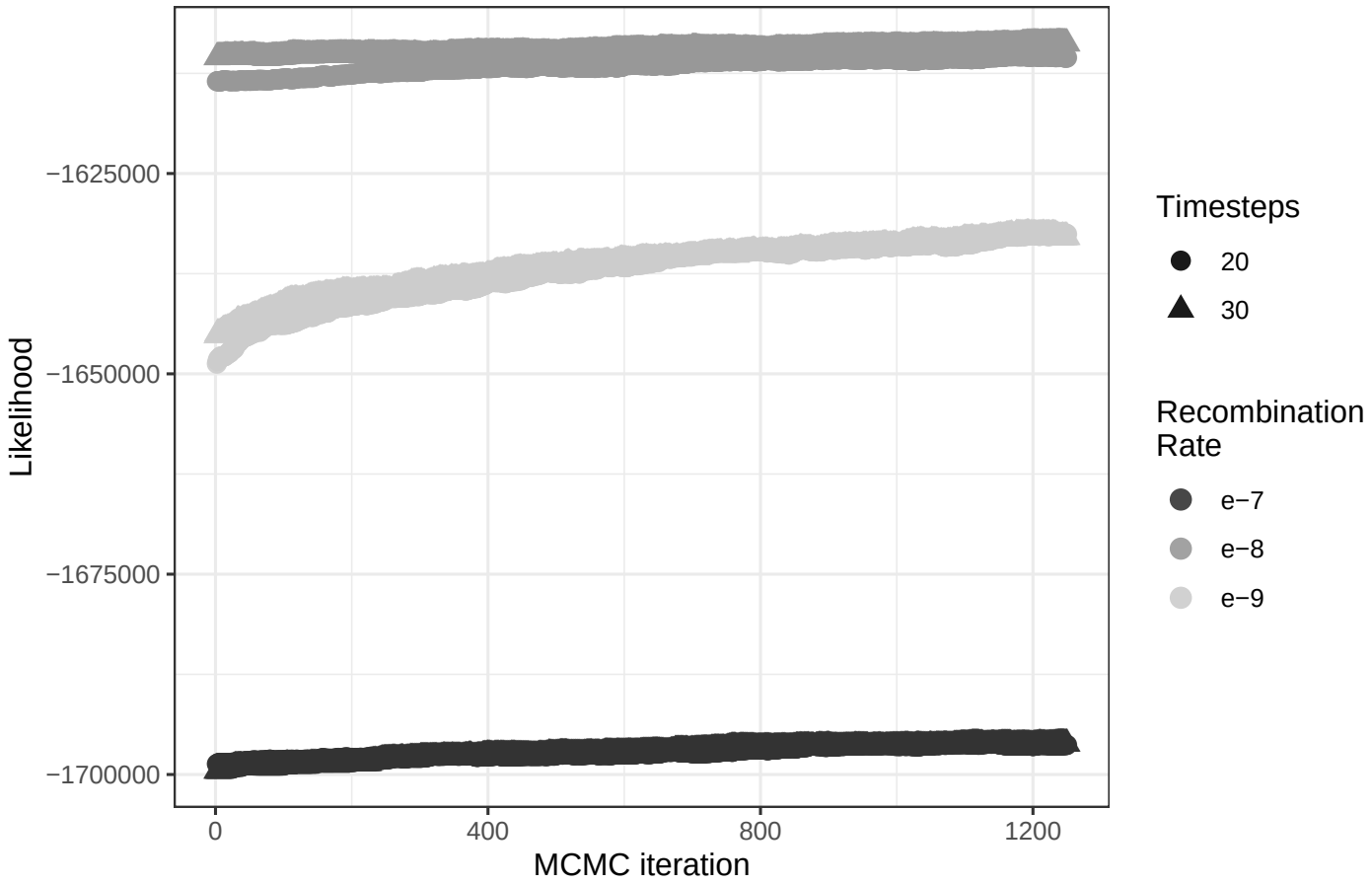

Figure 2 - figure supplement 2. The influence of both the number of timesteps coalescent events are estimated over (t) and constant recombination rate magnitude (r) on ARG likelihood across 1,250 MCMC iterations.

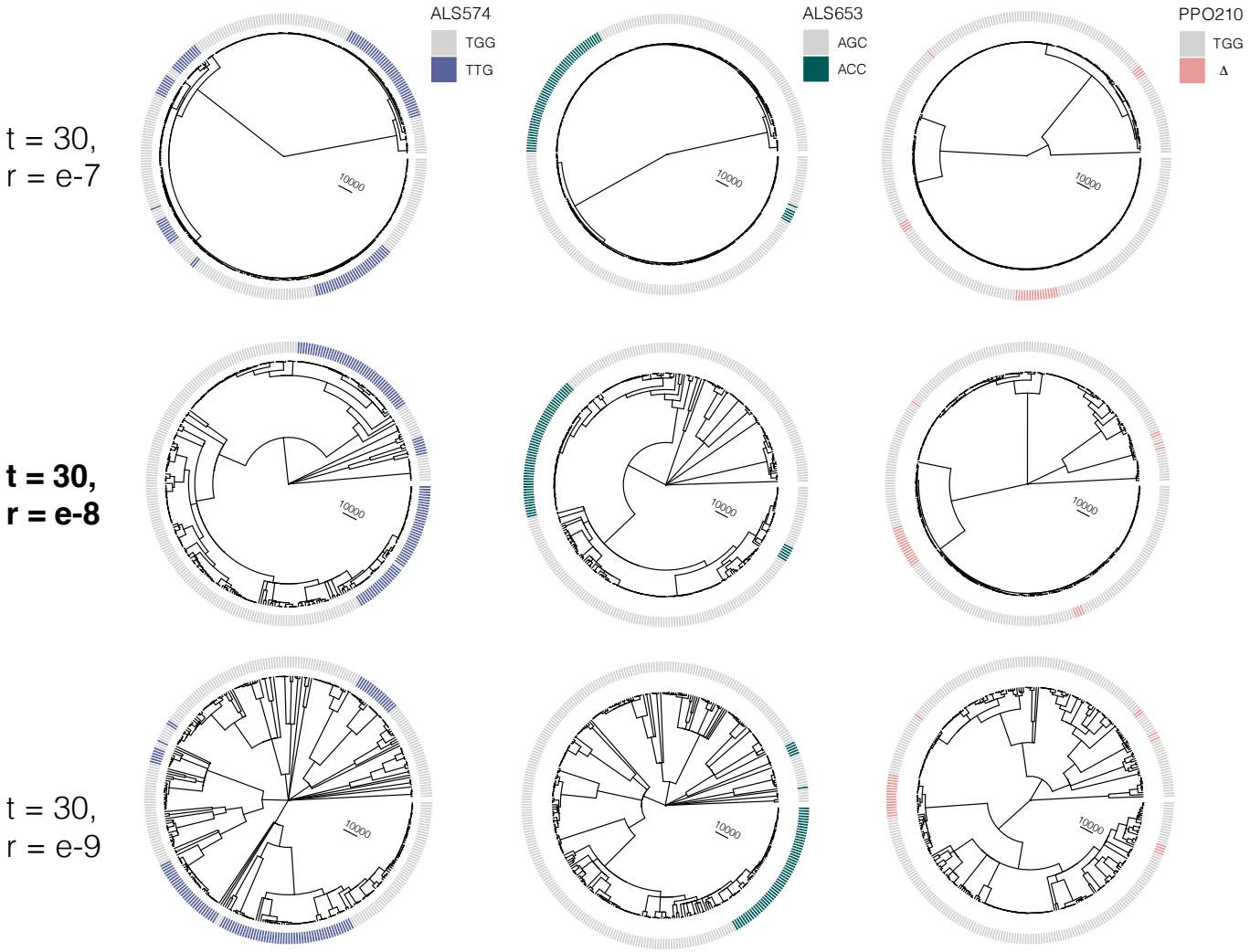

Figure 2 - figure supplement 3. At the most likely number of (timesteps = 30), the influence of increasing the recombination rate parameter constant value from r=e-7 to r=e-9 on the tree sequence inference from the most likely ARG.

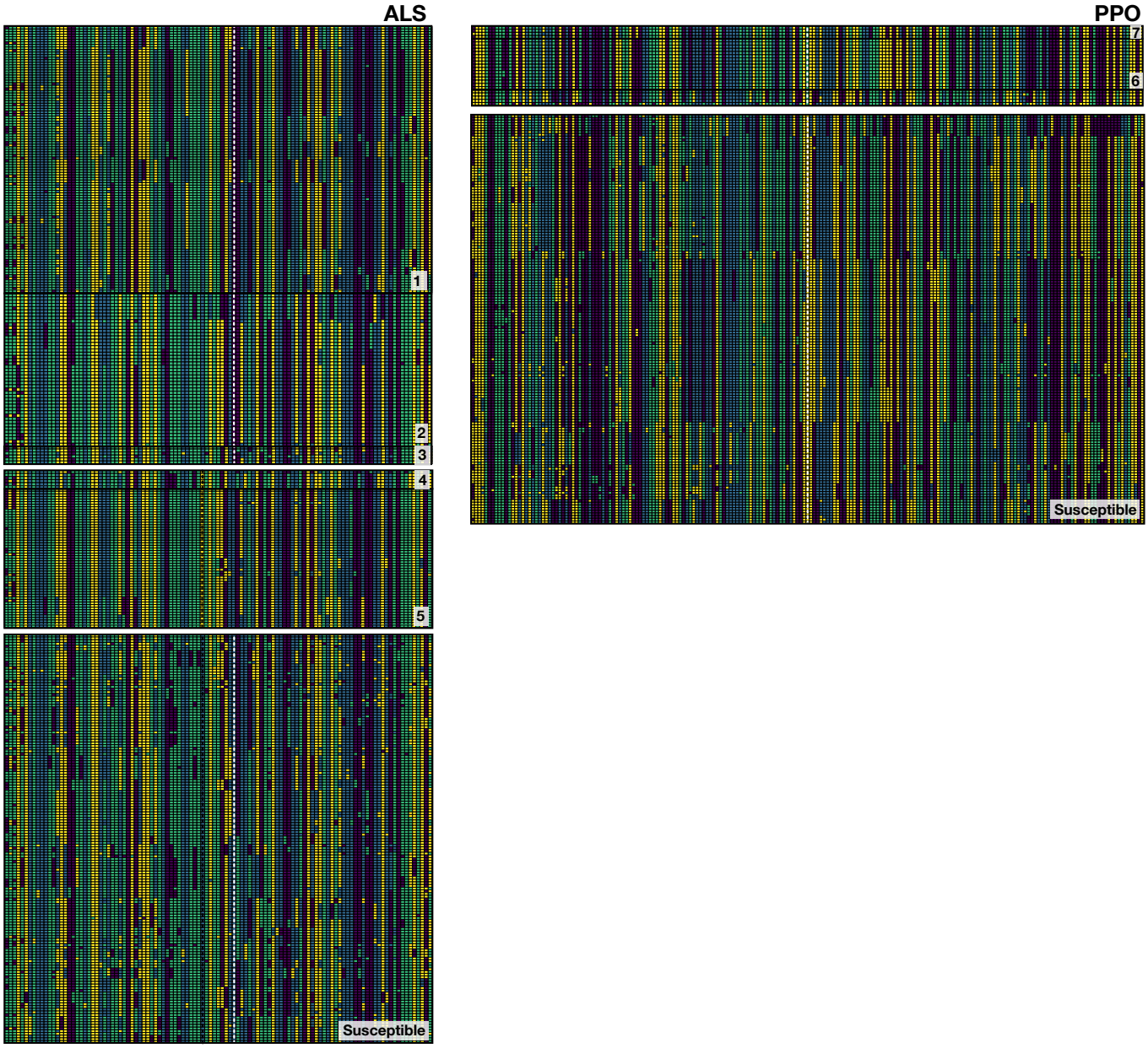

Figure 2 - figure supplement 4. Phased haplotypes corresponding to distinct origins of target-site resistance mutations at the ALS Trp-574-Leu, ALS Ser-653-Asn, and PPO ΔGly210 positions, relative to susceptible lineages. Haplotypes corresponding to each named origin in Figure 2 depicted by number on the bottom right of each set.

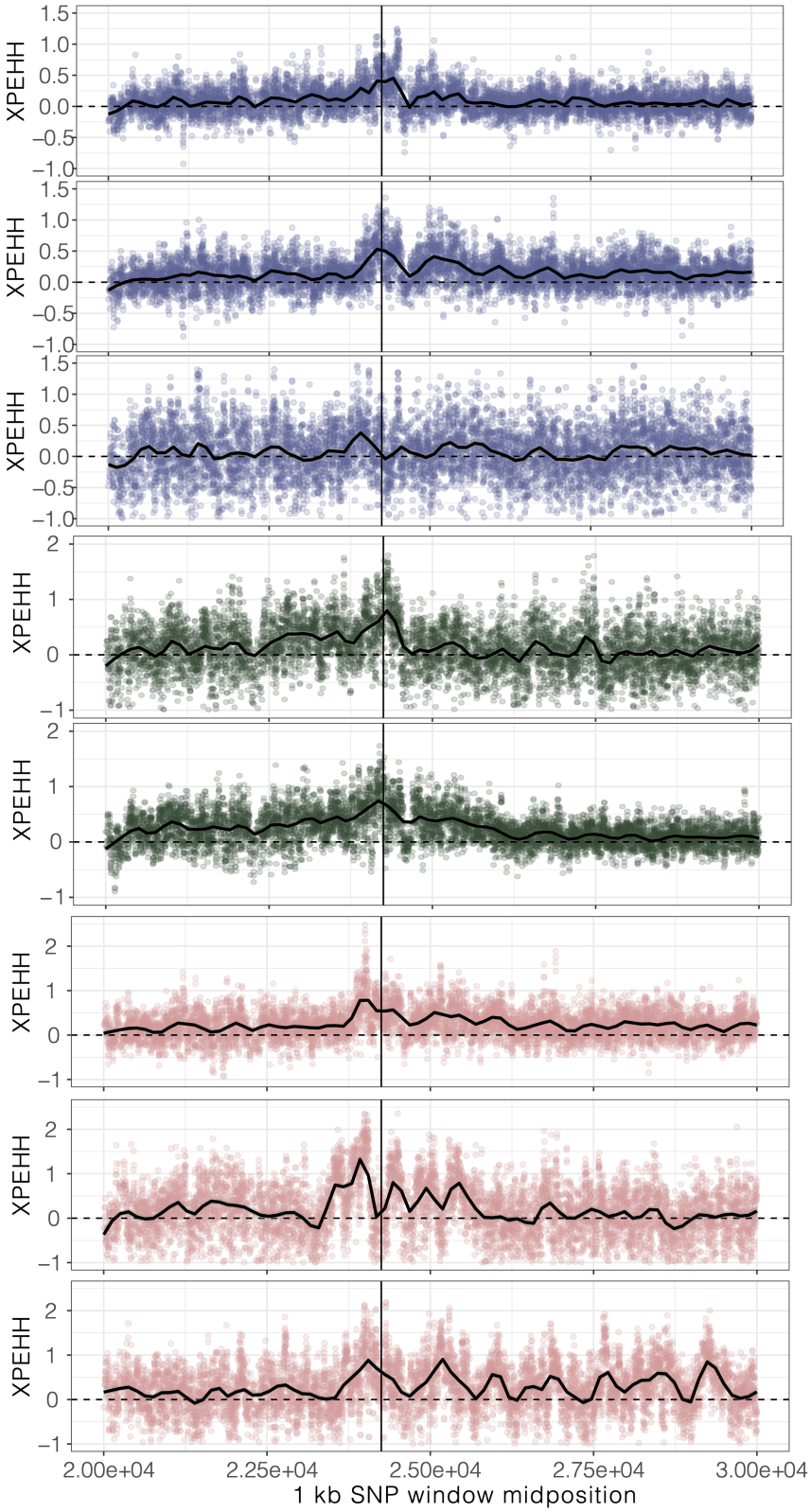

Figure 2 - figure supplement 5. XPEHH, cross-population extended haplotype homozygosity, for haplotypes mapping to each origin of ALS Trp-574-Leu, ALS Ser-653-Asn, and PPO ΔGly210 (1-7, top to bottom, plus unplaced PPO ΔGly210 haplotypes as the last row) in comparison to susceptible haplotypes

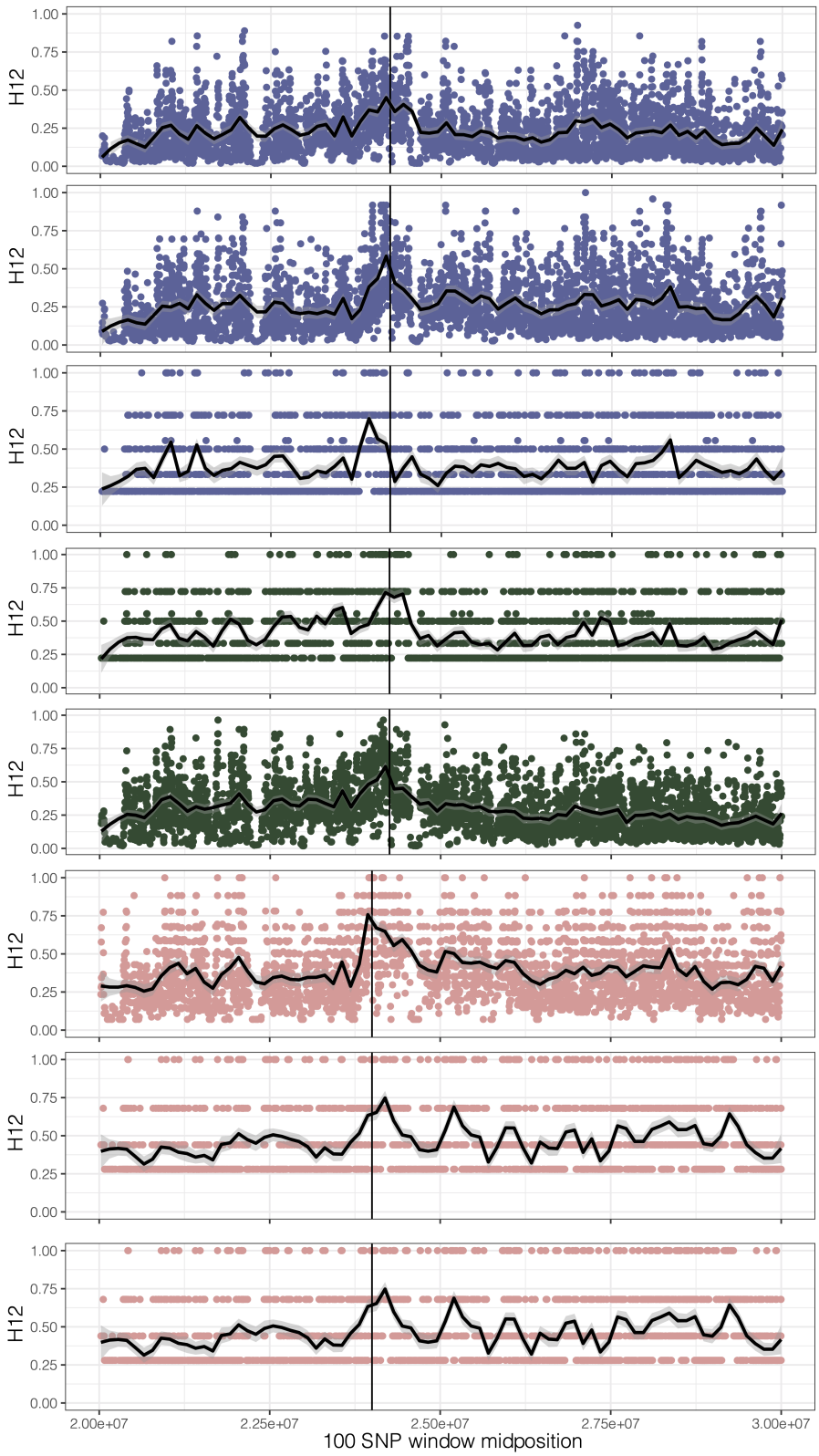

Figure 2 - figure supplement 6. H12, homozygosity of the two most common haplotypes, for each origin of ALS Trp-574-Leu, ALS Ser-653-Asn, and PPO ΔGly210 (1-7, top to bottom, plus unplaced PPO ΔGly210 haplotypes as the last row). A hamming distance of 5 was used, collapsing haplotypes that have fewer than 5 differences in each 100 SNP window.

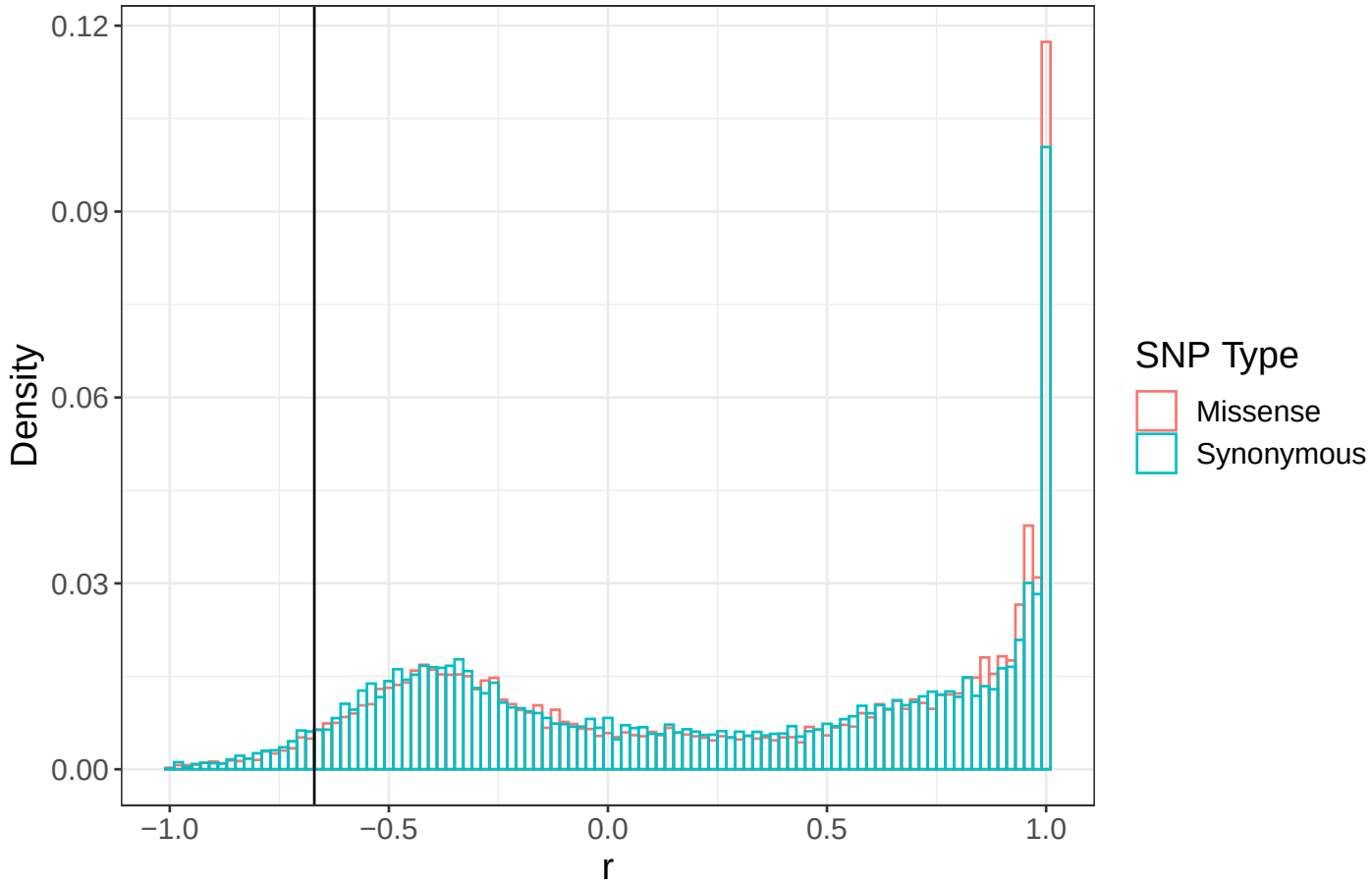

Figure 4 - figure supplement 1. Signed LD (*r*) between pairs of missense (nonsynonymous) and synonymous SNPs of similar frequency (Minor Allele Count > 20) and physical distance (<500 bp) as ALS Trp-574-Leu and ALS Ser-653-Asn across the genome. Observed signed LD between ALS Trp-574-Leu and ALS Ser-653-Asn in Essex indicated by the vertical line (*r*=-0.69).

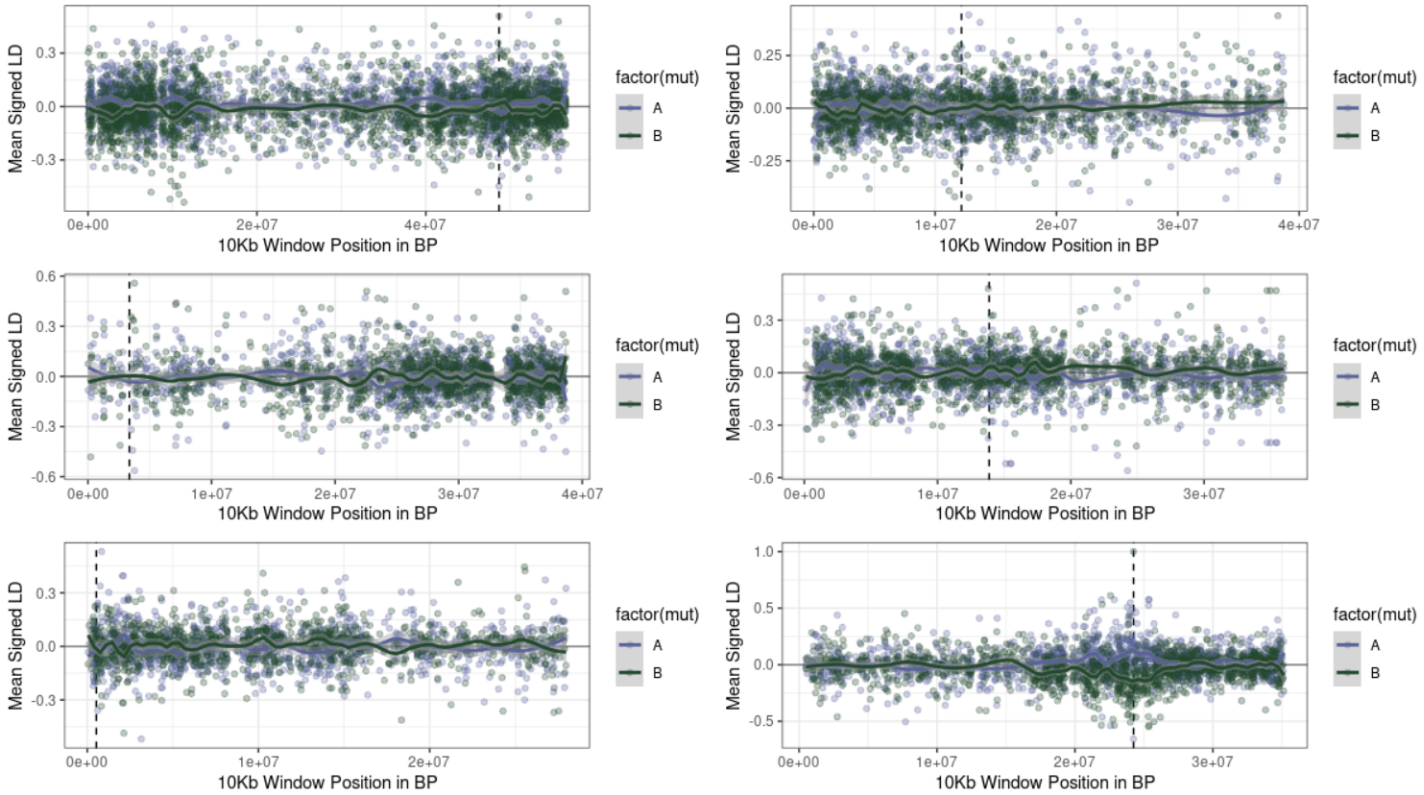

Figure 4 - figure supplement 2. Examples of the extent of extended repulsion elsewhere across the genome, for pairwise comparisons in Figure 4 - figure supplement 1 that were more extreme than observed. The first five figures show the limited breadth of repulsion associated with these pairwise comparisons, in contrast to the extended repulsion between ALS Trp-574-Leu and ALS Ser-653-Asn (bottom right)

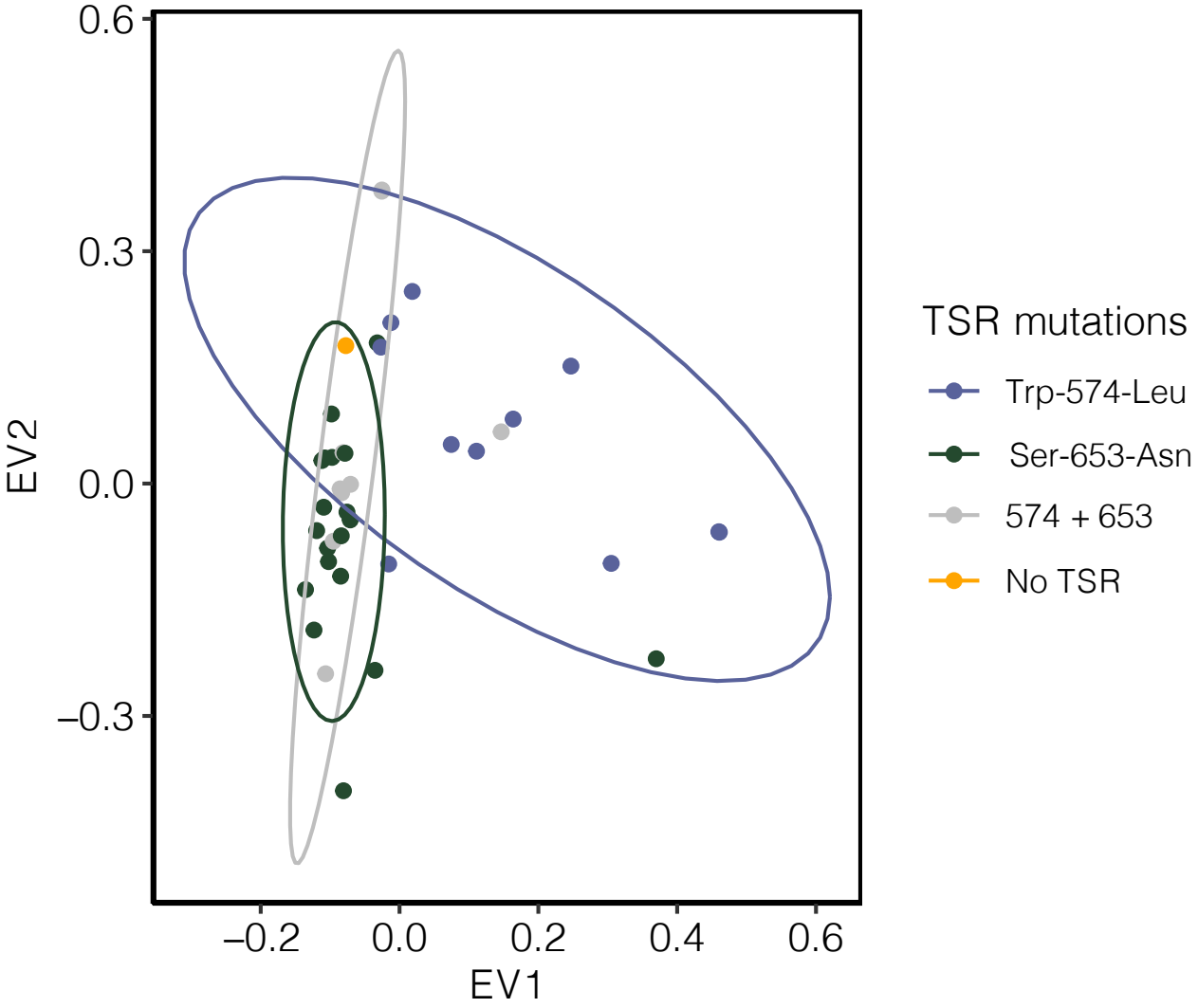

Figure 4 - figure supplement 3. Signature of population structure for 10 Mb around ALS based on PCAs of genotypes in Essex. Ellipses represent 95% CIs assuming a multivariate distribution.

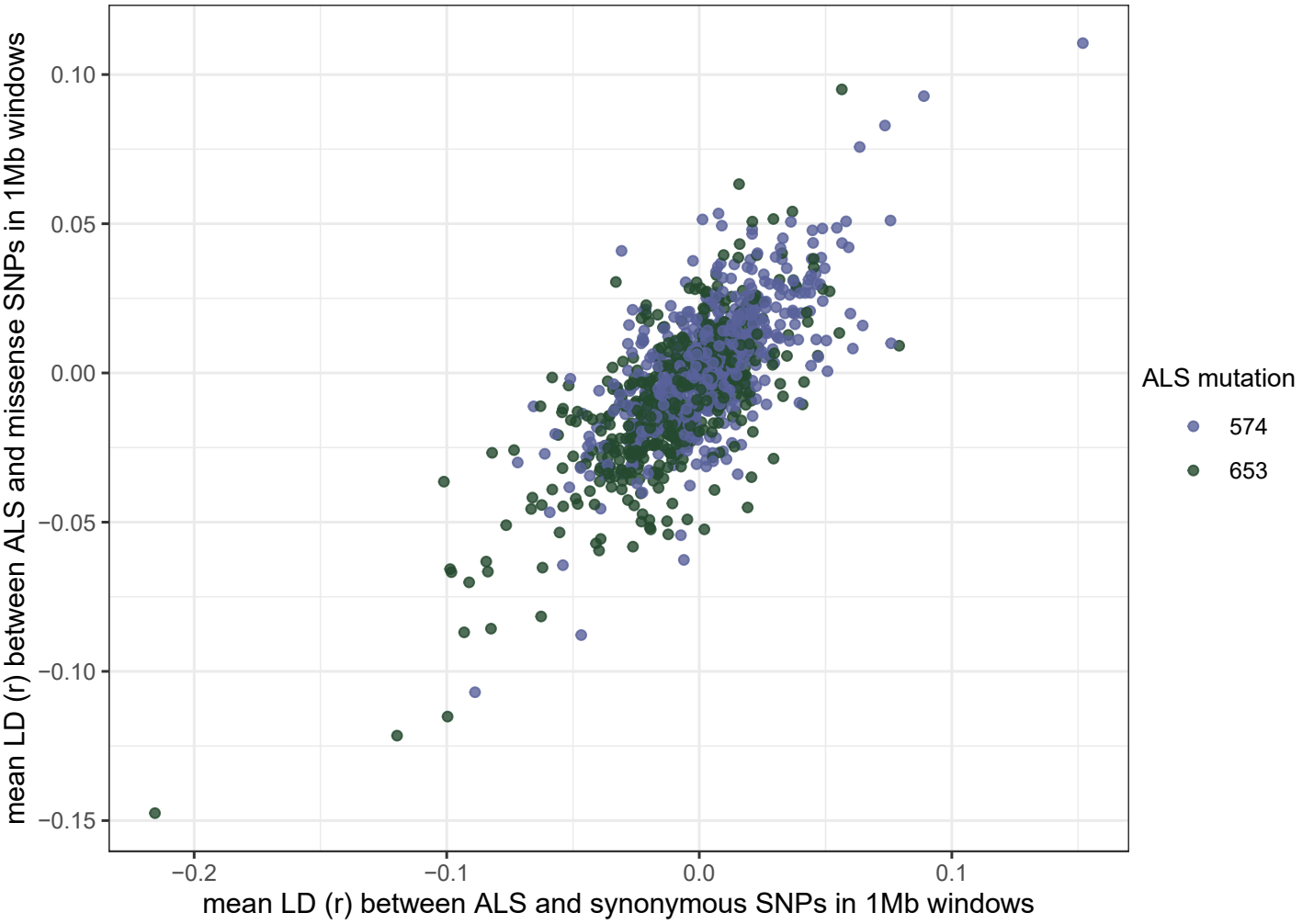

Figure 4 - figure supplement 4. Correlation of signed LD (*r*) between two target-site resistance mutations, and synonymous or missense mutations across the genome. Each point represents the mean r between all synonymous or missense mutations and an ALS mutation in a non-overlapping 1 Mb window.