Multiple modes of convergent adaptation in the spread of glyphosate-resistant *Amaranthus tuberculatus*

Julia M. Kreiner[a,1], Darci Ann Giacomini[b], Felix Bemm[c,2], Bridgit Waithaka[c], Julian Regalado[c], Christa Lanz[c], Julia Hildebrandt[c], Peter H. Sikkema[d], Patrick J. Tranel[b], Detlef Weigel[c,1], John R. Stinchcombe[a,e], and Stephen I. Wright[a,f]

[a]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON M5S 3B2, Canada; [b]Department of Crop Sciences, University of Illinois at Urbana–Champaign, Urbana, IL 61801; [c]Department of Molecular Biology, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany; [d]Department of Plant Agriculture, University of Guelph, Ridgetown, ON N0P 2C0, Canada; [e]Koffler Scientific Reserve, University of Toronto, King City, ON L7B 1K5, Canada; and [f]Centre for the Analysis of Genome Evolution and Function, University of Toronto, Toronto, ON M5S 3B2, Canada

Contributed by Detlef Weigel, August 26, 2019 (sent for review February 8, 2019; reviewed by Graham Coop and Alison F. Feder)

The selection pressure exerted by herbicides has led to the repeated evolution of herbicide resistance in weeds. The evolution of herbicide resistance on contemporary timescales in turn provides an outstanding opportunity to investigate key questions about the genetics of adaptation, in particular the relative importance of adaptation from new mutations, standing genetic variation, or geographic spread of adaptive alleles through gene flow. Glyphosate-resistant *Amaranthus tuberculatus* poses one of the most significant threats to crop yields in the Midwestern United States, with both agricultural populations and herbicide resistance only recently emerging in Canada. To understand the evolutionary mechanisms driving the spread of resistance, we sequenced and assembled the *A. tuberculatus* genome and investigated the origins and population genomics of 163 resequenced glyphosate-resistant and susceptible individuals from Canada and the United States. In Canada, we discovered multiple modes of convergent evolution: in one locality, resistance appears to have evolved through introductions of preadapted US genotypes, while in another, there is evidence for the independent evolution of resistance on genomic backgrounds that are historically nonagricultural. Moreover, resistance on these local, nonagricultural backgrounds appears to have occurred predominantly through the partial sweep of a single haplotype. In contrast, resistant haplotypes arising from the Midwestern United States show multiple amplification haplotypes segregating both between and within populations. Therefore, while the remarkable species-wide diversity of *A. tuberculatus* has facilitated geographic parallel adaptation of glyphosate resistance, more recently established agricultural populations are limited to adaptation in a more mutation-limited framework.

parallel evolution | herbicide resistance | de novo mutation | gene flow | population genomics

Glyphosate-resistant *Amaranthus tuberculatus* was first reported in Missouri in 2005 but has since been reported in 19 US states (1), with resistant biotypes harming corn and soybean yields (2, 3). Resistance to glyphosate in weed populations is widespread, likely as a result of the rapid adoption of and reliance on glyphosate weed control technology; 84% of soybeans, 60% of cotton, and 20% of corn grown in the United States by 2004 carried transgenes for glyphosate resistance, despite Roundup Ready technology—the combination of glyphosate weed control with transgenic glyphosate resistance—only having been introduced 8 y earlier (4). Agriculturally associated *A. tuberculatus* weed populations emerged in Canada in the province of Ontario only in the early 2000s, with glyphosate resistance following a decade later (3, 5). As with other herbicides, resistance in weed populations can evolve via substitutions at the direct target of glyphosate, 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), or by polygenic adaptation involving different loci in the genome (6–10). More often, glyphosate resistance in the genus *Amaranthus* has an unusual genetic basis: amplification of the *EPSPS* locus (11–15). Gene amplification apparently evolved independently in 2 *Amaranthus* species (14, 16, 17), raising the possibility that it could have evolved multiple times independently within a single species, or even population (18). While glyphosate resistance has been studied from multiple angles (15, 19–23), the recent discovery of glyphosate-resistant *A. tuberculatus* in southwestern Ontario affords the unique opportunity to evaluate the evolutionary origins of herbicide resistance, whether it has arisen through de novo mutation or standing genetic variation, and the role of gene flow in the recent spread of herbicide resistance in an agronomically important weed.

Native to North America, the dioecious, wind-pollinated *A. tuberculatus* has a history marked by the interaction of 2 lineages or subspecies [sensu Costea and Tardif (19) and Pratt and Clark (20)], thought to have been diverging on either side of the Mississippi River until they were brought back into contact through human-mediated disturbance (21, 22). Morphological, herbarium, and microsatellite evidence point to an expansion of

Significance

While evolution has been thought of as playing out over millions of years, adaptation to new environments can occur very rapidly, presenting us with key opportunities to understand evolutionary dynamics. One of the most amazing examples of real-time evolution comes from agriculture, where due to the intense use of a few herbicides, many plant species have evolved herbicide resistance to become aggressive weeds. An important question has been whether herbicide resistance arises only rarely and then spreads quickly, or whether herbicide resistance arises all the time de novo. Our work with glyphosate resistance in US Midwestern and Canadian populations of *Amaranthus tuberculatus* reveals the answer to be, "it depends," as we surprisingly find examples for both modes of evolution.

Author contributions: J.M.K., P.J.T., D.W., J.R.S., and S.I.W. designed research; J.M.K., B.W., C.L., J.H., and P.H.S. performed research; D.W. contributed new reagents/analytic tools; J.M.K., D.A.G., F.B., J.R., and S.I.W. analyzed data; and J.M.K., D.W., J.R.S., and S.I.W. wrote the paper.

Reviewers: G.C., University of California, Davis; and A.F.F., Stanford University.

The authors declare no competing interest.

Data deposition: Scripts and code are available at https://github.com/jkreinz/Amaranthus-population-genomics. The data reported in this paper have been deposited in the European Nucleotide Archive (ENA), https://www.ebi.ac.uk/ena/data/view/PRJEB31711 (project no. PRJEB31711).

[1]To whom correspondence may be addressed. Email: julia.kreiner@mail.utoronto.ca or weigel@weigelworld.org.

[2]Present address: KWS SE, 37555 Einbeck, Germany.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1900870116/-/DCSupplemental.

First published September 30, 2019.

the western var. *rudis* subspecies range limits over the last 50 y, while the range of more eastern var. *tuberculatus* subspecies is thought to be stagnant and constrained to riparian habitats (22, 23). With the timing of the var. *rudis* expansion coinciding with the invasion of *A. tuberculatus* into agricultural environments, var. *rudis* is hypothesized to be a predominant driver of this agricultural invasion (23).

Sequencing and Collections. We assembled a high-quality reference genome for *A. tuberculatus* from a single individual from 58 Gb (~87× genome coverage) long-read data collected on the Pacific Biosciences Sequel platform using 15 SMRT cells. After assembly, polishing, and haplotype merging, the reference genome consisted of 2,514 contigs with a total size of 663 Mb and an N50 of 1.7 Mb (see *SI Appendix*, Table S1, for details). Our final genome size is consistent with recent cytometric estimates of 676 Mb (SE = 27 Mb) for *A. tuberculatus* (24). The new reference included 88% of the near-universal single-copy orthologs present in BUSCO's Embryophyta benchmarking dataset, with 6% marked as duplicates (25). For chromosome-scale sweep scan analyses, we further scaffolded our contigs onto the fully resolved *Amaranthus hypochondriacus* genome (26), resulting in 16 final pseudomolecules (which included 99.8% of our original assembly; *Methods*) for population genetic analyses.

We resequenced whole genomes of 163 individuals to about 10× coverage from 19 agricultural fields in Missouri, Illinois, and 2 regions where glyphosate resistance has recently appeared in Ontario: Essex County, an agriculturally important region in southwestern Ontario; and Walpole Island, an expansive wetland and First Nation reserve with growing agricultural activity (Fig. 1*B*). Populations from the Midwest (Missouri and Illinois) had been previously assayed for glyphosate-resistant phenotypes, and from qPCR of genomic DNA it was found that resistance was predominantly conferred through *EPSPS* copy number amplification (11). Populations from Walpole Island and Essex County in Ontario were sampled in 2016 after reports from farmers that they were not controlled by herbicides. We also sampled 10 individuals from riparian habitats in Ontario near Walpole Island and Essex County, as a nonagricultural, natural Canadian comparison (Fig. 1*B*). Genome-wide diversity in *A. tuberculatus* is quite high, even relative to other wind-pollinated outcrossers (27), with neutral diversity (mean pairwise difference) at 4-fold degenerate sites being 0.041. The frequency of glyphosate resistance in the sampled agricultural fields ranged from 13 to 88%, based on greenhouse

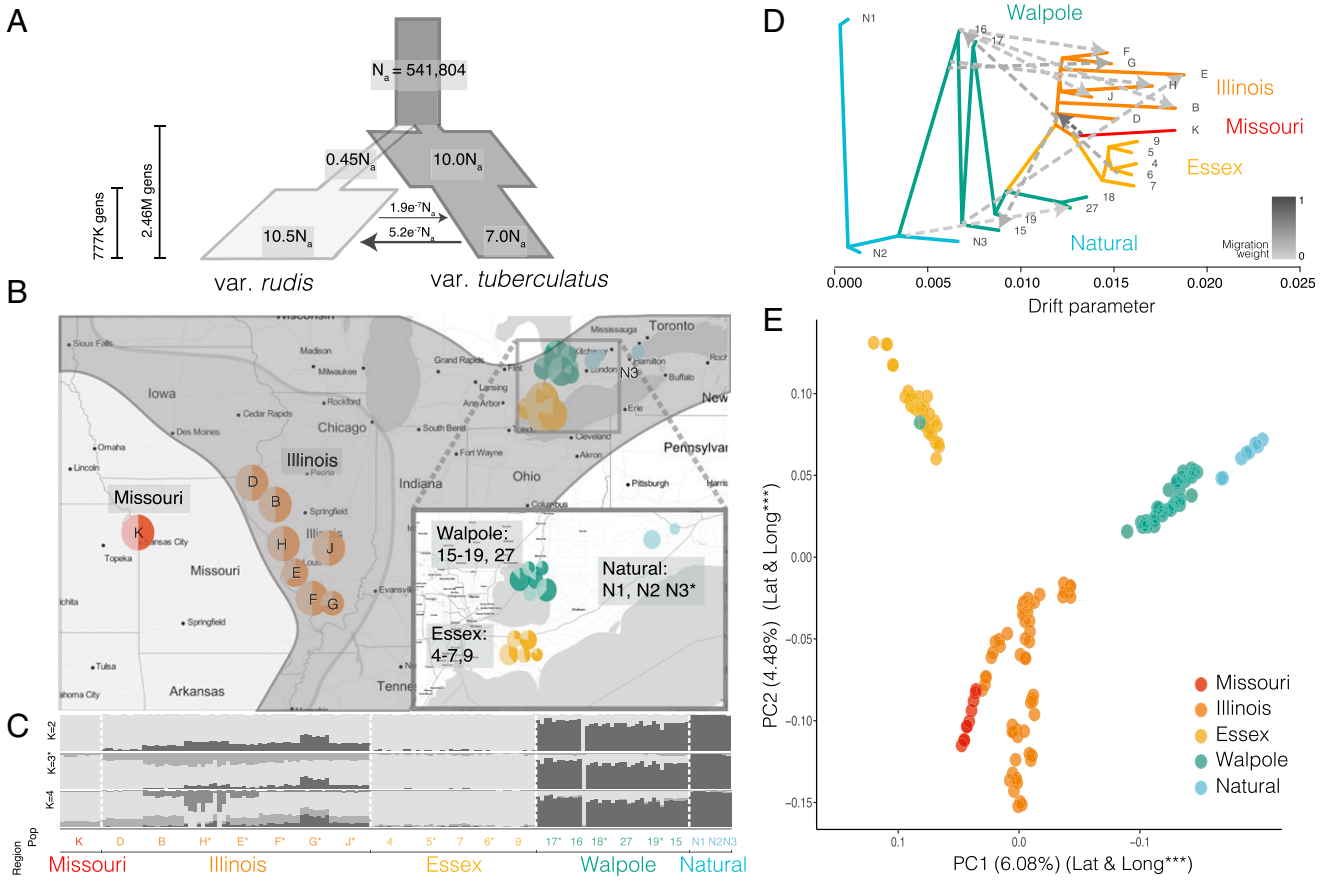

Fig. 1. Population structure and demographic history in *Amaranthus tuberculatus*. (*A*) Demographic model of the 2 *A. tuberculatus* subspecies (N_a = ancestral effective population size, with migration and other effective population size estimates scaled accordingly). (*B*) Geographic map of phenotyped and sequenced populations with variable glyphosate resistance in the United States and Canada. Natural populations are nonagricultural, Canadian populations without glyphosate resistance. Populations are color-coded by region (maintained throughout text). Dark portions of each pie chart indicate proportion of resistant individuals. *Inset* shows close-up of agricultural and natural Ontario populations. Historical range limits of the species are indicated on the map, with the same coloration as for the demographic model: *A. tuberculatus* var. *rudis* in light gray to the west, and *A. tuberculatus* var. *tuberculatus* in dark gray to the east [spp. range limits described by Sauer (21)]. (*C*) STRUCTURE plot of admixture across regions and populations from west (*Left*) to east (*Right*), with most likely number of clusters (K) = 3. The darkest ancestry group corresponds to var. *tuberculatus*, and the lightest corresponds to var. *rudis*. (*D*) *treemix* results showing the maximum likelihood of relatedness between populations based on allele frequencies, with population labels at the tip of each branch, dashed arrows indicating the amount (color) and direction of migration between populations, and the drift parameter reflecting the amount of genetic drift that has occurred between populations. (*E*) PCA of all individuals, with both PC1 and PC2 significantly relating to both longitude and latitude (*SI Appendix*, Table S2).

trials (Fig. 1*B* and *Methods*). Plants from natural populations in Ontario had no glyphosate resistance.

Demography of *Amaranthus tuberculatus*. To dissect the demographic context of convergent adaptation to glyphosate, we characterized genome-wide patterns of population structure, demography, and differentiation. Population structure, demographic modeling (Fig. 1), and phenotypic characterization confirmed the presence of the 2 previously hypothesized ancestral lineages, *A. tuberculatus* var. *rudis* and *A. tuberculatus* var. *tuberculatus* (22, 23). Population structure and investigations of the genome-wide proportion of introgression [*f* statistic (28)] largely reflected previous accounts of the historical range limits (22): Natural Ontario populations had the diagnostic indehiscent seed phenotype and were genetically homogeneous for ancestry of the var. *tuberculatus* lineage; Missouri samples were homogeneous for the var. *rudis* lineage; while Illinois, a region of sympatry in the historical range of the 2 subspecies, showed signs of introgression from var. *tuberculatus* {mixture proportion, *f* [95% confidence interval (CI)] = 0.1342 [0.126, 0.143], using Missouri as a reference (28, 29)} (Fig. 1 *B* and *C*). Genetic differentiation (F_{ST}) between individuals with ancestry homogenous for different lineages at $K = 2$ was 0.212, on par with or greater than that between other congeners (30). Moreover, both longitude and latitude significantly explained both PC1 and PC2 of the SNP matrix (*SI Appendix*, Table S2), with PC1 separating var. *rudis* and var. *tuberculatus* ancestry, and PC2 separating Canadian and American accessions (Fig. 1*E*). These patterns of diversity resulted in principal-component representation that, with few exceptions, reflected the geography of our samples. The most likely *tuberculatus-rudis* demographic model was one of secondary contact, with var. *rudis* having undergone a bottleneck followed by a dramatic expansion, which may be indicative of this subspecies' rapid colonization of agricultural fields across North America (Fig. 1*A*).

Demographic Origins of Canadian Agricultural Populations. Analyses of newly problematic agricultural populations in Ontario provides a unique angle for tracking the demographic source of the *A. tuberculatus* agricultural invasion. Populations from Essex County fell completely within the var. *rudis* cluster, with a *treemix* model indicating that Essex populations were derived from the most western Missouri population (Fig. 1*E*), the source of almost the entire Essex genome (*f*[95% CI] = 0.996 [0.985, 1]). Furthermore, while Essex grouped with Walpole and Natural populations on PC2, it was found at the other end on PC1, more different from Canadian populations than even the most geographically distant Missouri population (Fig. 1*B*). These patterns of population structure were distinct from the continuous gradient of southwest–northeast ancestry previously reported (23) and supports the hypothesis that glyphosate-resistant *A. tuberculatus* was introduced to Ontario from the United States through seed-contaminated agricultural machinery (3, 5) or animal-mediated seed dispersal (31).

In contrast to Essex as a likely introduction of a preadapted genotype to a new locale, populations from Walpole Island, where glyphosate resistance was first reported in Ontario (5), were mainly of the native, eastern var. *tuberculatus* type (Fig. 1). However, the convergent evolution of var. *tuberculatus* into agricultural fields may not be solely the result of de novo mutations. Populations from Walpole Island showed signs of introgression from var. *rudis* (*f*[95% CI] = 0.225 [0.215, 0.236]), while *treemix* indicated that Walpole may be a hot spot for gene flow, with 9 of 10 total migration events across the tree involving Walpole (explaining an additional 2.5% of SNP variation compared to a migration-free model) (Fig. 1*C*). Thus, both adaptive introgression from the western var. *rudis* clade and/or de novo adaptation from local natural populations could be playing a role in the evolution of resistance and adaptation to agricultural environments in Walpole.

Despite the considerable level of var. *rudis* introgression into Walpole, these populations were similarly differentiated from nearby natural populations homogenous for var. *tuberculatus* ancestry as they were from comparably admixed populations in Illinois [$F_{ST\ (Walpole\text{-}Nat)} = 0.0286$; $F_{ST\ (Walpole\text{-}Illinois)} = 0.0284$; *SI Appendix*, Fig. S1]. This, along with the tight clustering of Walpole and Natural populations in the PCA and structure analyses, implies that Walpole populations experienced strong and rapid local adaptation to agricultural environments upon its conversion from wetland, which may have been facilitated by introgression from var. *rudis*. We therefore sought to find genes that were highly differentiated between Walpole and Natural populations, putatively involved in agricultural adaptation. A Gene Ontology (GO) enrichment test for the top 1% of genes with excess differentiation between Walpole and natural populations identified genes with monooxygenase and oxidoreductase molecular

A

Molecular Function	Fold Enrichment	raw P value	FDR
alkane 1-monooxygenase activity	46.08	0.00000682	0.021
oxidoreductase activity	40.32	0.0000101	0.0156
Biological Process			
histone lysine methylation	16.8	2.50E-05	4.92E-02
peptidyl-lysine methylation	14.23	8.73E-06	5.16E-02
peptidyl-amino acid modification	3.95	1.15E-05	3.40E-02
Protein Class			
DNA methyltransferase	15.12	0.00153	0.0531
ribonucleoprotein	6.3	0.00158	0.0459
transfer/carrier protein	4.55	0.00119	0.0516
dehydrogenase	2.92	0.000743	0.0431
oxidoreductase	2.42	0.0000143	0.00249
hydrolase	1.78	0.00164	0.0409

B

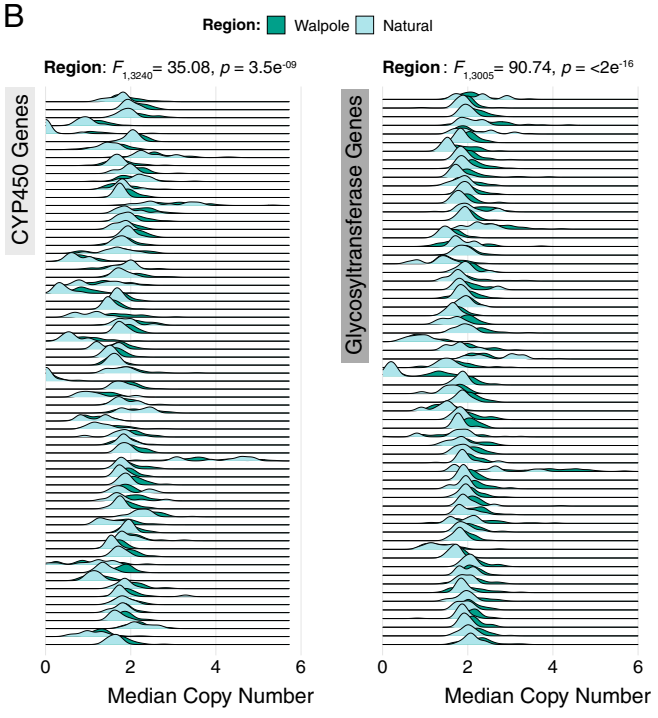

Fig. 2. Enrichment and expansion of CYP450 and glycosyltransferase gene families in the transition from natural to agricultural in Walpole. (*A*) GO categories that were significantly enriched in an analysis of the 99th percentile of F_{st} outliers between Walpole and Natural populations. Light gray indicates GO categories that include CYP450s; dark gray indicates the category that includes glycosyltransferases. (*B*) Evidence for copy number expansion of both CYP450 and glycosyltransferase genes in Walpole relative to Natural populations. Each row represents a single annotated gene of a given gene family, with each density plot representing the distribution of the median copy number, inferred from 100-bp windows, across individuals for that gene.

function, with histone methylation biological function, and several protein classes involved in transport and amino acid/protein modification (Fig. 2*A*). Of particular interest were 2 enriched GO categories important in metabolic, nontarget site resistance: oxidoreductases, which include cytochrome P450s, and peptidylamino acid modifiers, which include glycosyltransferases (7). CYP450s and glycosyltransferases work consecutively to detoxify herbicides in plant cells by catalyzing hydroxylation and glc-conjugation (32). In addition to these 2 gene families being highly differentiated between Walpole and Natural populations, we also investigated the possibility for copy number expansion or contraction. An analysis of the median copy number of 100-bp windows within each gene, for each individual (*Methods*), revealed that genes in both gene families consistently had significantly expanded copy number in Walpole populations (least square means: CYP450s, 1.82 [95% CI, 1.80, 1.85]; glycosyltransferases, 2.04 [95% CI, 2.02, 2.06]) compared to natural populations (least square means: CYP450s, 1.67 [95% CI, 1.62, 1.71]; glycosyltransferases, 1.83 [95% CI, 1.79, 1.86]) (Fig. 2*B*). Despite this widespread pattern across 69 CYP450 and 64 glycosyltransferase genes, no CYP450s or glycosyltransferases were significantly correlated with our phenotypic rating of glyphosate resistance after Holm's correction for multiple testing. A possible explanation is that the copy number expansion of these gene families confers resistance to herbicides other than just glyphosate, or, more broadly, is a result of the transition from natural to agricultural habitats.

Genetic Mechanisms of Glyphosate Resistance. Two major evolutionary paths to glyphosate resistance are amplification of wild-type *EPSPS* and nonsynonymous mutations in *EPSPS* that make the enzyme resistant to glyphosate inhibition. To better understand the genetic mechanisms underpinning glyphosate resistance, we investigated how variation in resistance relates to these 2 classes of *EPSPS* mutations. Using our genomic data to quantify copy number (*Methods*), we found that of 84 individuals assayed in the greenhouse as resistant (resistance, ≥2/5 rating; Fig. 3), 60 (71%) had elevated *EPSPS* copy number (>1.5; as in ref. 13). However, almost 26% (22 of 83) of individuals assayed as susceptible had an *EPSPS* copy number >1.5 (compared to 15% or 13 of 88 individuals for a >2 cutoff). Apart from errors in phenotyping or copy number estimation, this implies that intermediate copy number amplification alone may not always be sufficient for resistance, e.g., if amplified copies are not properly expressed. While *EPSPS* amplification was most frequent in the Midwest (83% [33 of 40] of resistant individuals, compared to 70% [16 of 23] in Walpole and 52% [11 of 21] in Essex), copy number in resistant individuals was on average almost twice as high in Walpole (~9 copies on average, compared to 5 in the Midwest and 4 in Essex). Previous estimates of *EPSPS* copy number in resistant *A. tuberculatus* were up to 17.5 copies relative to diploid susceptibles (11); we found 2 individuals in Walpole with an estimated 29 copies (Fig. 3). A regression of resistance onto copy number was significant in all 3 geographic regions (Walpole, P = 2.6e-07; Essex, P = 0.002; Midwest, P = 3.5e-06), explaining 48% of the variation in resistance in Walpole, but only 23% and 27% in Essex and the Midwest. In these latter 2 regions, however, an additional 10% of variation was explained by a nonsynonymous substitution at codon 106, the most common and well-characterized genetic mechanism of glyphosate resistance across species outside of the genus *Amaranthus* (1), in this instance causing a change from proline to serine (Fig. 3).

The presence of 2 types of target-site resistance mechanisms, copy number increase and nonsynonymous mutation at a critical codon, implies parallel evolution of the resistance phenotype through independent genetic pathways. While the well-known P106S nonsynonymous mutation can account for some of the resistance unexplained by copy number increase alone, other uncharacterized nontarget site mechanisms are likely contributing as well and thus providing a further path to convergent evolutionary outcomes. In addition to shedding light on the prevalence of different resistance mechanisms, the population genomic data allowed us to determine whether our most prevalent genetic mechanism, namely the *EPSPS* gene amplification, arose multiple times.

Genetic Origins of the *EPSPS* Amplification. Our chromosome-scale genome assembly provided a unique opportunity to determine the structure and genomic footprint of selection around the amplified *EPSPS* locus in different populations. Across all populations, copy number increase was not restricted to the 10-kb *EPSPS* gene—individuals identified to have increased copy number at *EPSPS* also had a correlated increase in the mean and variance of copy number for up to 6.5 Mb of the reference genome (23.5 to 30 Mb on chromosome 5), encompassing 108 genes (*SI Appendix*, Fig. S2).

Characterizing signals of selection for a high copy number region can be challenging. First, typical population genetic statistics ignore potential variation among gene copies that are collapsed into a single haplotype. Ideally, phasing of a multicopy region would allow for full resolution of the single-nucleotide polymorphism (SNP) differences within and between haplotypes. However, very recent gene amplification is expected to limit SNP variation among amplified copies, and will also hinder phasing approaches from short read data. Second, variation may not be recognized because of allelic dropout of low-copy variants. Analysis of the relationship between EPSPS copy number and homozygosity in our dataset suggested that higher-copy haplotypes did not feature more SNP variation than lower-copy haplotypes, implying that generally few new mutations distinguish among amplified copies (*SI Appendix*, Fig. S3). Nonetheless, to control for the possibility of residual SNP differences that exist among gene copies and/or for allelic dropout, we created a consensus haploid sequence by random downsampling to 1 allele per heterozygous site. Because we downsampled SNP by SNP, we do not expect our downsampling to be biased toward any particular haplotype.

While the *EPSPS*-related amplification showed the strongest selective signal on all of chromosome 5, we found distinct selective sweep patterns in the different agricultural regions. We

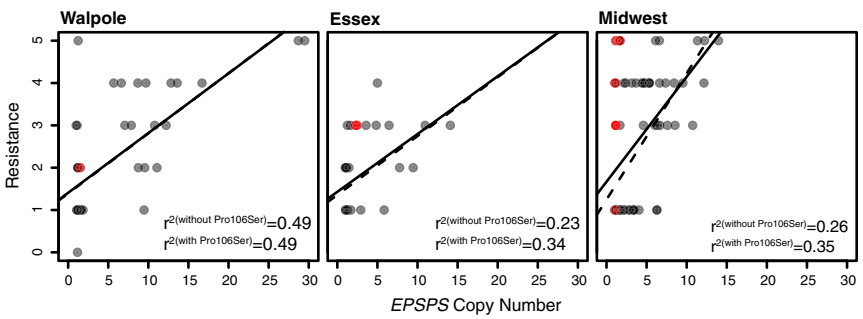

Fig. 3. *EPSPS* copy number variation among individuals, and its relationship with resistance. *EPSPS* copy number significantly explains phenotypic resistance within each agricultural region (solid linear regression line) (Walpole, P = 2.6e-07; Essex, P = 0.002; Midwest, P = 3.5e-06), with P106S substitution in *EPSPS* (red individuals) increasing explained resistance in Essex and the Midwest, but not in Walpole (dashed linear regression line).

ran *Sweepfinder2* (33, 34) across chromosome 5 to identify focal windows with a site-frequency spectrum particularly skewed by selection (while controlling for recombination) relative to genome-wide 4-fold degenerate sites. *Sweepfinder2* estimated the strongest amplification-related sweep signal in Walpole. In contrast to Essex or the Midwest, the top 1% of apparently selected 10-kb windows on chromosome 5 were localized to the amplified *EPSPS* region in Walpole (*SI Appendix*, Fig. S4). Moreover, there was a marked reduction in genetic diversity (mean pairwise differences) around *EPSPS*, as well as elevated differentiation (F_{st}) and extended haplotype homozygosity [XP-EHH score (35)] in Walpole individuals with the *EPSPS* amplification, implying a hard selective sweep, but not in Essex or Midwest individuals with increased *EPSPS* copy number (Fig. 4).

These differences in the extent of the amplification-related sweep signals across agricultural regions may be a consequence of how often *EPSPS* amplification has evolved; a hard sweep would be indicative of it having arisen only once, while soft sweeps would point to multiple origins (36–39). To investigate this further, we mapped *EPSPS* copy number onto a maximum-likelihood haplotype tree produced from SNP variants in *EPSPS*, and compared the phylogeny with phenotypic resistance and nonsynonymous target-site resistance status (Fig. 5). Indeed, the agricultural regions differed in the inferred number of independent copy number increases. Whereas there appears to have been only one amplification event in Walpole, Essex haplotypes of individuals with copy number increases are interspersed with susceptible haplotypes, both within and between populations. Similarly, haplotypes from Midwest individuals with *EPSPS* amplification are distributed across the gene tree, although some local populations show clustering indicative of a local hard sweep, implying independent evolutionary origins among populations and occasionally within populations in the Midwest (Fig. 5).

Together, these analyses suggest that gene amplification has occurred multiple times independently to different extents across the geographic range. However, it is possible that recombination and de novo mutation after amplification have contributed to the apparent soft sweep signal. To further test for multiple independent origins, we looked at the similarity in the copy number profiles of the *EPSPS* region, which should also be independent of any possible artifacts due to minority allele dropout in resequencing data. The copy number profiles of the amplified region varied considerably across our samples, and especially across agricultural regions (Fig. 6*A*), consistent with multiple independent amplification events. To quantify this, we calculated the Spearman's rank correlation coefficient of normalized sequence coverage in the 1-Mb chromosomal segment surrounding *EPSPS* between all possible pairs of individuals with copy number increases—this quantifies only the similarity in the rank, and not amplitude, of the landscape of copy number across loci within the segment (Fig. 6*B*). In agreement with our polymorphism-based inferences, the 2 Canadian regions showed very different patterns; coverages in different Walpole individuals were very highly correlated (average of Spearman's $\rho = 0.95$), suggesting the spread of a single amplification haplotype through a hard selective sweep. In contrast, there was much a lower average correlation across all Essex individuals region-wide ($\rho = 0.56$), and this was the case even when looking at the average within-population correlations rather than the single region-wide average (within-population average, e.g., $\rho = 0.54$ and 0.61), suggesting different haplotypes had independently experienced copy number increases (Fig. 6). Similar to Essex, there appeared to be multiple amplification haplotypes in the Midwest (average for all individuals, $\rho = 0.47$), but within-population correlations were higher, consistent with hard ($\rho = 0.94$, 0.95, 0.93) or soft sweeps ($\rho = 0.66$, 0.74, 0.75) (Fig. 6).

Discussion

The patterns of genetic differentiation and similarity in amplification profiles among agricultural regions helped us to distinguish between modes of adaptation, the evolutionary mechanisms by which glyphosate resistance has spread, and the extent of constraint on this particular genetic pathway. Although the Walpole population showed signs of admixture from var. *rudis*, Walpole individuals were clearly differentiated at *EPSPS* from both Essex and Midwest individuals (*SI Appendix*, Fig. S5). Moreover, copy number profiles were almost perfectly correlated within Walpole, but showed low correlations with Essex and the Midwestern individuals (Fig. 6*B*). This suggests that glyphosate resistance in Walpole evolved independently, likely from selection on a de novo

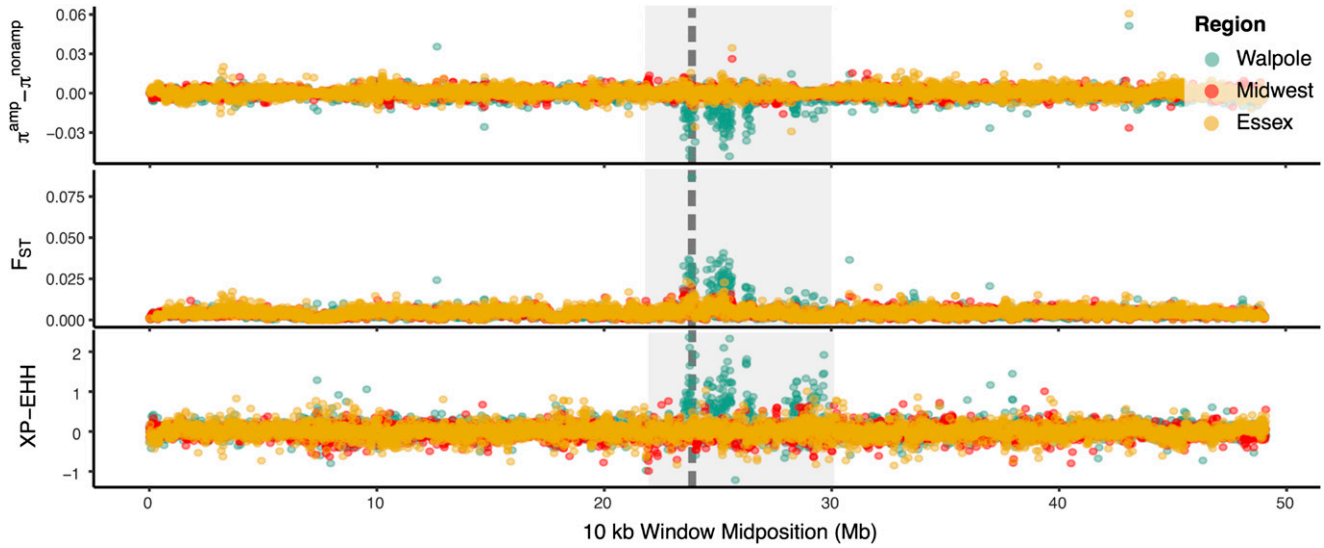

Fig. 4. Population genetic signals of selection related to copy number increase around *EPSPS* on chromosome 5. The deficit of diversity (*Top*), relative differentiation (*Middle*), and difference in extended haplotype homozygosity [XP-EHH (34)] (*Bottom*) is compared between amplified and nonamplified individuals in each agricultural region. *EPSPS* is delimited by the vertical gray dashed line, while the *EPSPS*-linked region undergoing amplification is shown by the light gray box, spanning 23.5 to 30 Mb on chromosome 5.

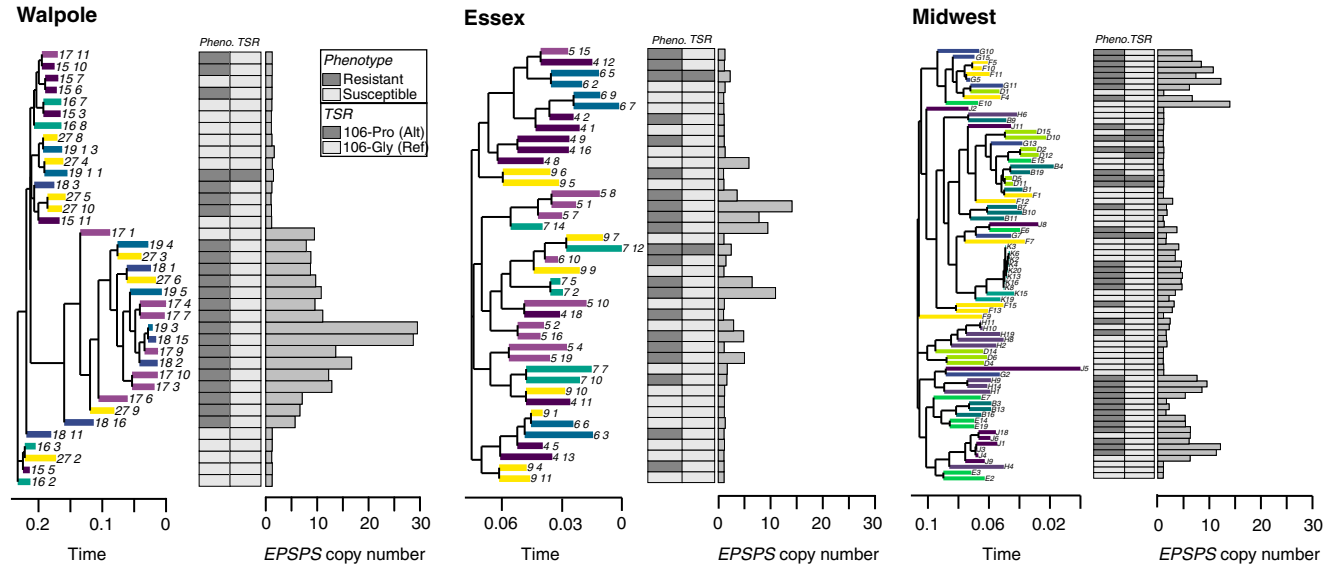

Fig. 5. Diversity of *EPSPS* amplification origins across agricultural regions. For each agricultural region, we show a haplotype tree of based on SNPs within *EPSPS* (with tips colored by population of origin), alongside a bar plot of *EPSPS* copy number, and a matrix of phenotypic resistance and target site-resistance status for the Gly-106-Pro mutation.

amplification event, although we do not know whether the amplification occurred in Walpole, or whether this allele was introgressed from an unsampled population. In Essex, the lack of within-region correlation in *EPSPS* copy profiles and sporadic high correlations with individuals from different Midwestern populations (Fig. 6*B*), suggest multiple independent amplification events. Together with the lack of genetic differentiation between Essex and Midwest (both genome-wide and on all of chromosome 5, including EPSPS; *SI Appendix*, Fig. S5), this suggests that Essex was either directly colonized by a diverse glyphosate-resistant population from the Midwest, or that a prior glyphosate-susceptible population in Essex was replaced by glyphosate-resistant individuals from the Midwest.

In summary, we have found multiple modes of convergent evolution underlying the spread of glyphosate resistance in North American *A. tuberculatus* populations. There is evidence for a single *EPSPS* amplification event that gave rise to the resistant populations in Walpole, distinct from amplification events in populations from another Canadian region, Essex County, and from populations in the US Midwest, where glyphosate resistance is older than in Canada. In contrast to the hard sweep in Walpole, glyphosate selection has left only soft selective sweep signals in the Midwest, because different haplotypes were amplified independently. Together with our analyses of population structure and demographic history, these results suggest that evolution on the more agriculturally naive, and recently bottlenecked *A. tuberculatus* var. *tuberculatus* background occurred in a mutation-limited framework, relying on evolutionary rescue via de novo mutation. In contrast, multiple independent amplification haplotypes have been maintained both within and among populations of *A. tuberculatus* var. *rudis*, likely resulting from its recent population expansion, long-range gene flow (as in Essex), and a longer history of spatially and temporally fluctuating selection [as suggested in Kreiner et al. (40)]. Therefore, demographic history and duration of selection have interacted to determine whether adaptation remains constrained to a mutation-limited framework.

A practical outcome of this work is that it informs on the scale of management that is needed to control herbicide resistance. Specifically, we suggest that with glyphosate resistance spreading across the range through seed translocation and independent adaptation, management efforts should be broadened to encompass both regional seed containment and local integrative control of herbicide-resistant weeds. We are faced with an additional challenge—that historically nonweedy lineages can adapt to an agricultural environment on rapid, contemporary timescales—calling for more consideration of how to prevent seemingly benign weeds from becoming problematic.

Methods

Plant Collections. Seeds were collected from Midwestern populations in 2010 (11), and from Ontario natural populations and agricultural fields in the fall of 2016. Agricultural fields in which *A. tuberculatus* appeared to be poorly controlled were sampled, biasing the collection toward populations with high levels of glyphosate resistance. These do not necessarily represent levels of resistance in a random sample.

High–Molecular-Weight DNA Extraction. High–molecular-weight (HMW) DNA was extracted from the leaf tissue of a single 28-d-old female *A. tuberculatus* plant from the Midwest using a modified version of the Doyle and Doyle nuclei isolation protocol (41). Nuclei isolation was carried out by incubating 30 g of ground leaf tissue in a buffer comprising tris(hydroxymethyl)aminomethane, potassium chloride, EDTA, sucrose, spermidine, and spermine tetrahydrochloride (Sigma-Aldrich). The homogenate was subsequently filtered using miracloth and precipitated by centrifugation. G2 lysis buffer, RNase A, and Proteinase K (Qiagen) were then added prior to an overnight incubation at 50 °C, followed by centrifugation at 4 °C. The supernatant containing the DNA solution was added to an equilibrated Qiagen genomic tip 100 (Qiagen). Genomic DNA was eluted and precipitated using isopropanol. Finally, HMW DNA was isolated by spooling.

SMRTbell Library Preparation and Sequencing. HMW genomic DNA was sheared to 30 kb using a Megaruptor 2 instrument (Diagenode SA). DNA damage and end repair were carried out prior to blunt adaptor ligation and exonuclease purification using ExoIII and ExoVII, in accordance with the protocol supplied by Pacific Biosciences (P/N 101-024-600-02; Pacific Biosciences). The resultant SMRTbell templates were size-selected using a BluePippin (SageScience) instrument with a 15-kb cutoff and a 0.75% DF Marker S1 high-pass 15- to 20-kb gel cassette. The final library was sequenced on a Sequel System (Pacific Biosciences) with v2 chemistry, MagBead loading, and SMRT Link UI v4 analysis.

Lucigen PCR-Free Library Preparation and Sequencing. Genomic DNA was fragmented to 350-bp size using a Covaris S2 Focused Ultrasonicator (Covaris). Subsequent end-repair, A-tailing, Lucigen adaptor ligation, and size selection

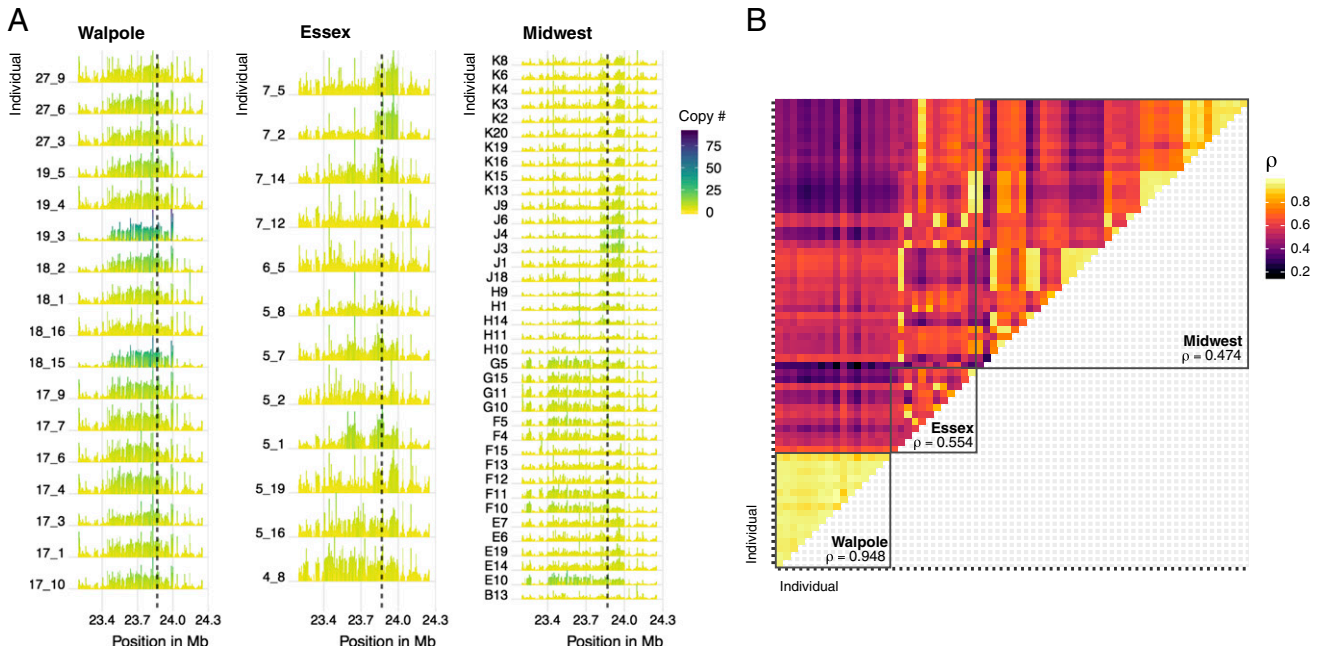

Fig. 6. Similarity of the *EPSPS* amplification inferred from copy number variation around the *EPSPS* gene. (*A*) The profile of normalized sequence coverage for 1 Mb around EPSPS (locus delimited by dashed line). Color of each bar under the curve represents copy number at that genomic position, where height of each profile has been rescaled by the max value within each region for clarity (Walpole, Essex, and Midwest, separately). Only individuals with at least 1.5× coverage relative to susceptible individuals were considered. (*B*) Pairwise Spearman's correlation coefficients over the same 1-Mb *EPSPS* region for individuals shown in *A*. Mean coefficient across all individuals within each agricultural region indicated. Order of individuals within each region as in *A*, from *Bottom* to *Top* and *Right* to *Left*.

were performed using the Lucigen NxSeq AMPFree Low DNA Library Kit (Lucigen). Libraries were quantified using a Qubit 2.0 instrument (Life Technologies), and library profiles were analyzed using a Bioanalyzer High Sensitivity Chip on an Agilent Bioanalyzer 2100 (Agilent Technologies). The libraries were sequenced to a coverage depth of 10× on an HiSeq 3000 instrument (Illumina) using a HiSeq 3000/4000 SBS kit and paired-end 150 base read chemistry. Raw fastq files were deposited to ENA (project no. PRJEB31711) (42).

Genome Assembly and Haplotype Merging. The genome was assembled from 58 Gb of Sequel long read data using *Canu* (version 1.6; genomeSize = 544 m; other parameters default) (43). Raw contigs were polished with *Arrow* (ConsensusCore2, version 3.0.0; consensus models S/P2-C2 and S/P2-C2/5.0; other parameters default) and *Pilon* (version 1.22; parameters default) (44). Polished contigs were repeat masked using *WindowMasker* (version 1.0.0; -checkdup; other parameters default) (45). Repeat-masked contigs were screened for misjoints and subjected to haplotype merging using *HaploMerger2* [commit 95f8589; identity = 80, other parameters default (46)]. A custom scoring matrix was supplied to both *lastz* steps of *Haplomerger2* (misjoint and haplotype detection). The scoring matrix was inferred from an all-vs.-all contig alignment using *minimap2* (version 2.10; preset asm10; other parameters default) (47) taking only the best contig-to-contig alignments into account. The final assembly was finished against the chromosome-resolved *A. hypochondriacus* genome (26) using *reveal finish* (commit 98d3ad1; –fixedgapsize –gapsize 15,000; other parameters default) (48). The 16 resulting pseudo-chromosomes represented 99.6% of the original assembly.

Alignment, SNP Calling, and Gene Annotation. We used *freebayes* (49) to call SNPs jointly on all samples. For whole-genome analyses, we used a thoroughly filtered SNP set following established guidelines (50, 51) adapted for whole-genome data: sites were removed based on missing data (>80%), complexity, indels, allelic bias (<0.25 and >0.75), whether there was a discrepancy in paired status of reads supporting reference or alternate alleles, and mapping quality (QUAL < 30, representing sites with greater than a 1/1,000 error rate). Individuals with excess missing data (>5%) were dropped. This led to a final, high-confidence SNP set of 10,280,132 sites. For *EPSPS*-specific analyses and genome-wide investigations that required invariant sites, we recalled SNPs with *samtools* (V1.7; ref. 52) and *bwa-mem* (V0.7.17; ref. 53). For this SNP set, sites were minimally filtered on mapping quality and missing data (keeping only sites with MQ >30 and <20% missing data), so that diversity estimates were not biased by preferentially retaining invariant or variant sites. For both SNP sets, we used *bwa-mem* to map to our fastqs to the reference genome. Bam files were sorted and duplicates marked with *sambamba* (V0.6.6; ref. 54), while cigars were split and read groups added with *picard* (V2.17.11).

We performed gene annotation on both our final assembly and the *A. hypochondriacus*-finished pseudoassembly using the *MAKER* pipeline (55). *A. tuberculatus*-specific repeats were identified using *RepeatModeler* (v1.0.11; ref. 13), combined with the RepBase repeat library, and masked with *RepeatMasker* (v4.0.7; ref. 14). This repeat-masked genome was then run through *MAKER* (v2.31.8), using expressed sequence tag evidence from an *A. tuberculatus* transcriptome assembly (56) and protein homology evidence from *A. hypochondriacus* (57). The gene models were further annotated using *InterProScan* (v69.0; ref. 58), resulting in a total of 30,771 genes and 40,766 transcripts with a mean transcript length of 1,245 bp. The mean annotation edit distance (AED) score was 0.21, and 98.1% of the gene predictions had an AED score of <0.5, indicating high-quality annotations.

Phenotyping. Seedlings from each population were grown in a 1:1:1:1 soil: peat:Torpedo Sand:LC1 (SunGro commercial potting mix) medium supplemented with 13-13-13 Osmocote in a greenhouse that was maintained at 28/22 °C day/night temperatures for a 16:8-h photoperiod. Plants were sprayed at the 5 to 7 leaf stage with 1,260 g of glyphosate (WeatherMax 4.5 L; Monsanto) per hectare. Fourteen days after treatment, plants were rated visually on a scale of 0 (highly sensitive) to 5 (no injury). Plants rated 2 or higher were classified as resistant. Prior to herbicide treatment, single leaf samples were taken from each plant and stored at −80 °C until ready for genomic DNA extraction. Tissue from plants rated as highly glyphosate resistant or susceptible were selected from each population for genomic DNA extraction using a modified cetrimonium bromide method (41).

Copy Number Estimates. Scaled coverage and copy number at *EPSPS* was estimated by dividing the coverage at each site across the focal region by the mode of genome-wide coverage after excluding centromeric regions (which have repeats and thus often abnormally high coverage) and regions of low coverage (<3×, indicative of technical coverage bias), which should represent the coverage of single-copy genes.

Structure, Demographic Modeling, and Summary Statistics. To model neutral demographic history and estimate neutral diversity, we used a Python script (available at https://github.com/tvkent/Degeneracy) to score 0-fold and 4-fold

degenerate sites across the genome. This procedure estimated 17,454,116 0-fold and 4,316,850 4-fold sites across the genome, and after intersecting with our final high-quality *freebayes*-called SNP set, resulted in 345,543 0-fold SNPs and 326,459 4-fold SNPs. The latter was used as input for demographic modeling.

Our 2-population demographic model of *A. tuberculatus* modeled the split between the *A. tuberculatus* var. *tuberculatus* and var. *rudis* subspecies by collapsing individuals into 1 of the 2 populations based on predominant ancestry as identified in our *STRUCTURE* analyses, estimated in ∂a∂i (V1.7.0) (59) using the pipeline available on https://github.com/dportik/dadi_pipeline (60). 1D and 2D site frequency spectrums were estimated using the program *easySFS* (https://github.com/isaacovercast/easySFS), and samples were projected downward to maximize the number of loci without missing data vs. number of individuals retained. We ensured that the log-likelihood of our parameter set had optimized by iterating the analysis over 4 rounds of increasing reps, from 10 to 40. We tested a set of 20 diversification models, with variation in split times, symmetry of migration, constancy of migration, population sizes, and size changes. The most likely inferred demography followed a model of secondary contact, where initially populations split without gene flow, followed by population size change with asymmetrical gene flow, and included 8 parameters: size of population 1 after split (nu1a), size of population 2 after split (nu2a), the scaled time between the split and the secondary contact (in units of $2*N_a$ generations) (T1), the scaled time between the secondary contact and present (T_2), size of population 1 after time interval (nu1b), size of population 2 after time interval (nu2b), migration from population 2 to population 1 ($2*N_a*m12$), and migration from population 1 to population 2 (m21). N_e was calculated by substituting the per-site θ estimate (after controlling for the effective sequence length to account for losses in the alignment and missed or filtered calls) and the Arabidopsis *thaliana* mutation rate (7×10^{-9}) (61) into the equation $\theta = 4Ne\mu$.

We used *PLINK* (V1.9; ref. 62) to perform a PCA of genotypes from our final *freebayes* SNP set after thinning to reduce the effects of sites that are in linkage disequilibrium, used STRUCTURE (V2.3.4) (63) to estimate admixture across populations, and *treemix* (V3) (64) to infer patterns of population splitting and migration events. To calculate summary statistics (π, F_{ST}, D_{xy}), we used scripts from the genomics general pipeline available at https://github.com/simonhmartin/genomics_general, binning SNPs into 100-kb windows with a step size of 10 kb. To estimate the proportion of introgression of var. *rudis* ancestry into Walpole agricultural populations in these genomic windows, we used the *f* statistic (but with nonoverlapping windows) (28). For investigation of introgression of Natural populations into Illinois (var. *tuberculatus* into var. *rudis*), we used Missouri as the reference ingroup. For investigation of introgression of Essex populations into Walpole (var. *tuberculatus* into var. *rudis*), we used Natural populations as the reference ingroup. Last, for investigation of introgression of Natural populations into Essex (var. *tuberculatus* into var. *rudis*), we used Missouri as the reference ingroup. To get CIs for the *f* statistic estimates, we performed jackknifing by calculating pseudovalues by removing one 250-kb block at a time.

For the outlier analysis of putative genes underlying contemporary agricultural adaptation in Walpole, we analyzed genome-wide differentiation (F_{st}) in 10-kb windows, and classified windows as outliers when they were in the top 1% for extreme differentiation. A GO enrichment test was then performed for these outlier regions, after finding their intersecting annotated *A. tuberculatus* genes, and their orthologs in *A. thaliana* using *orthofinder* (65). To look at the possibility of gene expansion in these enriched gene families, we first characterized normalized copy number in 100-bp windows within each annotated gene in that family, for every individual. We then characterized the median copy number across windows within each gene for each individual, as heterogeneous mapping of paralogs/orthologs due to differential levels of degeneration should lead to variation in copy number across windows within the gene. We then compared the distribution of the median copy number between Walpole and Natural population individuals for every gene. We tested whether the distribution of median copy number of each gene differed consistently across all genes and between Walpole and Natural populations by performing an ANOVA of region and gene ID, and allowing for an interaction. Scripts and code are available at https://github.com/jkreinz/Amaranthus-population-genomics (66).

Detecting Selective Sweeps and Estimating Recombination Rate. To detect differences in the strength and breadth of sweep signal associated with selection from glyphosate across geographic regions, we used SNPs called from the pseudoassembly of our *A. tuberculatus* reference. Sweep detection can be strongly influenced by heterogeneity in recombination rate, and so as a control (in our *Sweepfinder2* and *XPEHH* analyses), we used the interval function in *LDhat* (67) to estimate variable recombination rate independently across all 16 chromosomes of the pseudoassembly, using a precomputed lookup table for a θ of 0.01 for 192 chromosomes. Accordingly, we randomly subsetted individuals to retain only 96 individuals for computation of recombination rate estimates, which was implemented by segmenting the genome into 2,000 SNP windows, following the workflow outline in https://github.com/QuentinRougemont/LDhat_workflow.

To account for the fact that high-copy number loci may allow for increased diversity relative to single-copy regions, we randomly sampled 1 allele per locus along the length of chromosome 5 to create pseudohaploid haplotypes for our sweep scans. This ensures that any increased differentiation is due to differences among individuals, rather than among haplotypes within individuals. The *XP-EHH* scan (35), calculated based on the difference in haplotype homozygosity between amplified and nonamplified individuals for each geographic region after controlling for recombination rate, was implemented in *selscan* (68). Scripts available at https://github.com/simonhmartin/genomics_general were used for calculating differentiation and the difference in diversity. Pseudohaploid haplotypes were also used to calculate a maximum-likelihood tree for the 235 SNPs in *EPSPS*. For each tree, we realigned sequences before bootstrapping 1,000 replicates of our haplotree with *clustal omega* (69). In contrast to haplotype-based methods that required phased data, we also ran *Sweepfinder2* (33, 34), a program that compares the likelihood of a selective skew in the site frequency spectrum (SFS) at focal windows compared to the background SFS while controlling for heterogeneity in recombination rate. The SFSs of 10-kb windows across chromosome 5 were compared to the genome-wide SFSs at 4-fold degenerate sites, that for this analysis, was also randomly sampled for 1 allele per locus, for an equivalent comparison of the SFS. Last, we investigated similarity in the EPSPS amplification within and among populations and regions by estimating the Spearman's rank correlation coefficient for all pairwise comparisons of resistant, amplification-containing individuals. This was done for the 1-Mb region surrounding EPSPS.

ACKNOWLEDGMENTS. We thank Tyler Kent and Anna O'Brien for useful discussion; Rebecca Schwab, Fernando Rabanal, and Talia Karasov for comments; and Yunchen Gong for computational support. We also thank and recognize the Ojibwe, Potawatomi, and Odawa peoples of the Walpole Island First Nation. This work was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants (S.I.W. and J.R.S.); an NSERC E.W.R. Steacie Memorial Fellowship (S.I.W.); an NSERC Postgraduate Scholarship–Doctoral (J.M.K.); the Department of Ecology and Evolutionary Biology at University of Toronto (J.M.K.); International Max Planck Research School "Molecules to Organisms" (B.W.); Max Planck Society and Ministry of Science, Research and the Arts of Baden-Württemberg in the Regio-Research-Alliance "Yield Stability in Dynamic Environments" (D.W.).

1. I. Heap, "Herbicide resistant weeds" in *Integrated Pest Management*, Peshin, R., Pimentel, D., Eds. (Springer, Dordrecht, 2014), pp. 281–301.
2. L. E. Steckel, C. L. Sprague, Common waterhemp (*Amaranthus rudis*) interference in corn. *Weed Sci.* **52**, 359–364 (2004).
3. M. Costea, S. E. Weaver, F. J. Tardif, The biology of invasive alien plants in Canada. 3. *Amaranthus tuberculatus* (Moq.) Sauer var. *rudis* (Sauer) Costea & Tardif. *Can. J. Plant Sci.* **85**, 507–522 (2005).
4. G. M. Dill, Glyphosate-resistant crops: History, status and future. *Pest. Manag. Sci.* **61**, 219–224 (2005).
5. M. G. Schryver *et al*., Glyphosate-resistant waterhemp (*Amaranthus tuberculatus* var. *rudis*) in Ontario, Canada. *Can. J. Plant Sci.* **97**, 1057–1067 (2017).
6. M. D. Devine, A. Shukla, Altered target sites as a mechanism of herbicide resistance. *Crop Prot.* **19**, 881–889 (2000).
7. J. S. Yuan, P. J. Tranel, C. N. Stewart, Jr, Non-target-site herbicide resistance: A family business. *Trends Plant Sci.* **12**, 6–13 (2007).
8. C. Délye, M. Jasieniuk, V. Le Corre, Deciphering the evolution of herbicide resistance in weeds. *Trends Genet.* **29**, 649–658 (2013).
9. J. Guo *et al*., Nontarget-site resistance to ALS inhibitors in waterhemp (*Amaranthus tuberculatus*). *Weed Sci.* **63**, 399–407 (2015).
10. V. K. Nandula, J. D. Ray, D. N. Ribeiro, Z. Pan, K. N. Reddy, Glyphosate resistance in tall waterhemp (*Amaranthus tuberculatus*) from Mississippi is due to both altered target-site and nontarget-site mechanisms. *Weed Sci.* **61**, 374–383 (2013).
11. L. A. Chatham *et al*., EPSPS gene amplification is present in the majority of glyphosate-resistant illinois waterhemp (*Amaranthus tuberculatus*) populations. *Weed Technol.* **29**, 48–55 (2015).
12. L. Lorentz *et al*., Characterization of glyphosate resistance in *Amaranthus tuberculatus* populations. *J. Agric. Food Chem.* **62**, 8134–8142 (2014).
13. L. A. Chatham *et al*., A multistate study of the association between glyphosate resistance and EPSPS gene amplification in waterhemp (*Amaranthus tuberculatus*). *Weed Sci.* **63**, 569–577 (2015).

14. D.-H. Koo *et al*., Extrachromosomal circular DNA-based amplification and transmission of herbicide resistance in crop weed *Amaranthus palmeri*. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 3332–3337 (2018).
15. D.-H. Koo *et al*., Gene duplication and aneuploidy trigger rapid evolution of herbicide resistance in common waterhemp. *Plant Physiol.* **176**, 1932–1938 (2018).
16. T. A. Gaines *et al*., Gene amplification confers glyphosate resistance in *Amaranthus palmeri*. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1029–1034 (2010).
17. A. Dillon *et al*., Physical mapping of amplified copies of the 5-enolpyruvylshikimate-3-phosphate synthase gene in glyphosate-resistant *Amaranthus tuberculatus*. *Plant Physiol.* **173**, 1226–1234 (2017).
18. E. L. Patterson, D. J. Pettinga, K. Ravet, P. Neve, T. A. Gaines, Glyphosate resistance and EPSPS gene duplication: Convergent evolution in multiple plant species. *J. Hered.* **109**, 117–125 (2018).
19. M. Costea, F. J. Tardif, Conspectus and notes on the genus *Amaranthus* in Canada. *Rhodora* **105**, 260–281 (2003).
20. D. B. Pratt, L. G. Clark, *Amaranthus rudis* and *A. tuberculatus*, one species or two? *J. Torrey Bot. Soc.* **128**, 282–296 (2001).
21. J. Sauer, Revision of the dioecious amaranths. *Madrono* **13**, 5–46 (1955).
22. J. Sauer, Recent migration and evolution of the dioecious amaranths. *Evolution* **11**, 11–31 (1957).
23. K. E. Waselkov, K. M. Olsen, Population genetics and origin of the native North American agricultural weed waterhemp (*Amaranthus tuberculatus*; Amaranthaceae). *Am. J. Bot.* **101**, 1726–1736 (2014).
24. M. G. Stetter, K. J. Schmid, Analysis of phylogenetic relationships and genome size evolution of the *Amaranthus* genus using GBS indicates the ancestors of an ancient crop. *Mol. Phylogenet. Evol.* **109**, 80–92 (2017).
25. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
26. D. J. Lightfoot *et al*., Single-molecule sequencing and Hi-C-based proximity-guided assembly of *Amaranth* (*Amaranthus hypochondriacus*) chromosomes provide insights into genome evolution. *BMC Biol.* **15**, 74 (2017).
27. J. Chen, S. Glémin, M. Lascoux, Genetic diversity and the efficacy of purifying selection across plant and animal species. *Mol. Biol. Evol.* **34**, 1417–1428 (2017).
28. S. H. Martin, J. W. Davey, C. D. Jiggins, Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* **32**, 244–257 (2015).
29. R. E. Green *et al*., A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
30. M. G. Stetter, M. Vidal-Villarejo, K. J. Schmid, Convergent seed color adaptation during repeated domestication of an ancient new world grain. bioRxiv:10.1101/547943 (13 February 2019).
31. J. A. Farmer, E. B. Webb, R. A. Pierce , 2nd, K. W. Bradley, Evaluating the potential for weed seed dispersal based on waterfowl consumption and seed viability. *Pest Manag. Sci.* **73**, 2592–2603 (2017).
32. Q. Yu, S. Powles, Metabolism-based herbicide resistance and cross-resistance in crop weeds: A threat to herbicide sustainability and global crop production. *Plant Physiol.* **166**, 1106–1118 (2014).
33. M. DeGiorgio, C. D. Huber, M. J. Hubisz, I. Hellmann, R. Nielsen, SweepFinder2: Increased sensitivity, robustness and flexibility. *Bioinformatics* **32**, 1895–1897 (2016).
34. C. D. Huber, M. DeGiorgio, I. Hellmann, R. Nielsen, Detecting recent selective sweeps while controlling for mutation rate and background selection. *Mol. Ecol.* **25**, 142–156 (2016).
35. P. C. Sabeti *et al*.; International HapMap Consortium, Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).
36. J. Hermisson, P. S. Pennings, Soft sweeps: Molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**, 2335–2352 (2005).
37. P. S. Pennings, J. Hermisson, Soft sweeps II—molecular population genetics of adaptation from recurrent mutation or migration. *Mol. Biol. Evol.* **23**, 1076–1084 (2006).
38. P. S. Pennings, J. Hermisson, Soft sweeps III: The signature of positive selection from recurrent mutation. *PLoS Genet.* **2**, e186 (2006).
39. J. Hermisson, P. S. Pennings, Soft sweeps and beyond: Understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods Ecol. Evol.* **8**, 700–716 (2017).
40. J. M. Kreiner, J. R. Stinchcombe, S. I. Wright, Population genomics of herbicide resistance: Adaptation via evolutionary rescue. *Annu. Rev. Plant Biol.* **69**, 611–635 (2018).
41. J. Doyle, J. L. Doyle, Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochem. Bull.* **19**, 11–15 (1987).
42. J. M. Kreiner *et al*., Sequence data from PNAS 2019. European Nucleotide Archive. https://www.ebi.ac.uk/ena/data/view/PRJEB31711. Deposited 14 March 2019.
43. S. Koren *et al*., Canu: Scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
44. B. J. Walker *et al*., Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963 (2014).
45. A. Morgulis, E. M. Gertz, A. A. Schäffer, R. Agarwala, WindowMasker: Window-based masker for sequenced genomes. *Bioinformatics* **22**, 134–141 (2006).
46. S. Huang, M. Kang, A. Xu, HaploMerger2: Rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. *Bioinformatics* **33**, 2577–2579 (2017).
47. H. Li, Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
48. J. Linthorst, M. Hulsman, H. Holstege, M. Reinders, Scalable multi whole-genome alignment using recursive exact matching. bioRxiv:10.1101/022715 (17 July 2015).
49. E. Garrison, G. Marth, Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907 (20 July 2012).
50. H. Fang, Towards better understanding of artifacts in variant calling from high-coverage samples. http://repository.cshl.edu/30521/1/JC_April2014.pdf (2014). Accessed 11 January 2018.
51. J. B. Puritz, C. M. Hollenbeck, J. R. Gold, dDocent: A RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ* **2**, e431 (2014).
52. H. Li *et al*.; 1000 Genome Project Data Processing Subgroup, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
53. H. Li, (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 (26 May 2013).
54. A. Tarasov, A. J. Vilella, E. Cuppen, I. J. Nijman, P. Prins, Sambamba: Fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).
55. B. L. Cantarel *et al*., MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196 (2008).
56. C. W. Riggins, Y. Peng, C. N. Stewart , Jr, P. J. Tranel, Characterization of de novo transcriptome for waterhemp (*Amaranthus tuberculatus*) using GS-FLX 454 pyrosequencing and its application for studies of herbicide target-site genes. *Pest Manag. Sci.* **66**, 1042–1052 (2010).
57. J. W. Clouse *et al*., The *Amaranth* genome: Genome, transcriptome, and physical map assembly. *Plant Genome* **9**, 1 (2016).
58. P. Jones *et al*., InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
59. R. N. Gutenkunst, R. D. Hernandez, S. H. Williamson, C. D. Bustamante, Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5**, e1000695 (2009).
60. D. M. Portik *et al*., Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Mol. Ecol.* **26**, 5245–5263 (2017).
61. S. Ossowski *et al*., The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* **327**, 92–94 (2010).
62. S. Purcell *et al*., PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
63. J. K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
64. J. K. Pickrell, J. K. Pritchard, Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
65. D. M. Emms, S. Kelly, OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
66. J. M. Kreiner *et al*., Scripts from PNAS 2019. GitHub. https://github.com/jkreinz/Amaranthus-population-genomics. Accessed 19 September 2019.
67. A. Auton, G. McVean, Recombination rate estimation in the presence of hotspots. *Genome Res.* **17**, 1219–1227 (2007).
68. Z. A. Szpiech, R. D. Hernandez, Selscan: An efficient multithreaded program to perform EHH-based scans for positive selection. *Mol. Biol. Evol.* **31**, 2824–2827 (2014).
69. F. Sievers *et al*., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).

RESEARCH ARTICLE

Repeated origins, widespread gene flow, and allelic interactions of target-site herbicide resistance mutations

Julia M Kreiner[1]*‡§, George Sandler[1], Aaron J Stern[2], Patrick J Tranel[3], Detlef Weigel[4], John R Stinchcombe[1†], Stephen I Wright[1†]

[1]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada; [2]Graduate Group in Computational Biology, University of California, Berkeley, Berkeley, United States; [3]Department of Crop Sciences, University of Illinois Urbana-Champaign, Urbana, United States; [4]Department of Molecular Biology, Max Planck Institute for Biology Tübingen, Tübingen, Germany

Abstract Causal mutations and their frequency in agricultural fields are well-characterized for herbicide resistance. However, we still lack 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 19 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 extreme 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 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. Two such forces are 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.

Editor's evaluation

This paper studies the evolution of herbicide resistance in *Amaranthus tuberculatus*, a widespread agricultural weed. By illuminating how adaptive mutations arose and spread in this remarkable example of rapid human-induced adaptation, the study will be of interest to a broad audience, ranging from plant biologists interested in herbicide resistance to evolutionary biologists and population geneticists studying the fundamental factors and processes that govern rapid adaptation. The paper applies innovative population genetic methodology to support its primary finding that resistance mutations have evolved multiple times in parallel.

***For correspondence:** julia.kreiner@ubc.ca

[†]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

Competing interest: See page 20

Funding: See page 20

Preprinted: 10 May 2021
Received: 11 May 2021
Accepted: 16 January 2022
Published: 17 January 2022

Reviewing Editor: Philipp W Messer, Cornell University, United States

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, 2014**) and barring the probably minor role of interspecific hybridization. For acetolactate synthase (ALS) inhibiting herbicides alone, over 160 species have evolved resistance since the first report in 1986, which was only five years after their initial introduction (**Comai and Stalker, 1986**; **Heap, 2014**; **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, 2014**; **Kreiner et al., 2018**). In addition to repeated resistance evolution through distinct causal resistance loci, it is likely that for a single locus, resistance mutations have arisen repeatedly within a species (**Kreiner et al., 2019**). While these observations suggest herbicide resistance may be among the most extreme cases of contemporary parallel evolution 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 (**Küpper et al., 2018**), reconstructed complex genomic regions associated with resistance (**Molin et al., 2017**), and investigated patterns of selection on and the extent of convergence between loci conferring non-target site resistance (**Kreiner et al., 2020**; **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**; **Feder et al., 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 common waterhemp (*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 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 can arise (**Costea et al., 2005**; **Kreiner et al., 2018**) and an effective dispersal system (**Liu et al., 2017**). 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 (U.S. Department of Agriculture, National Agricultural Statistics Service (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 (*Dayan et al., 2018*; *Tranel, 2021*; *Tranel and Wright, 2002*; *Zhao et al., 2020*).

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*), 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.

Results

Types of target-site 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, potentially overrepresenting the frequency of resistance across the landscape.

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 - the target of a class of herbicides that inhibits electron transfer). Across 152 individuals, we found segregating variation at 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.

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 number of individuals 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 number 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, Ontario, Canada	0	12 (0.162)	12 (0.162)	1 (0.0135)	0	0	0	0	0
TSR mutations in Essex, Ontario, Canada	1 (0.0125)	23 (0.288)	35 (0.438)	0	1 (0.0125)	0	1 (0.0125)	2 (0.025)	3 (0.0375)
TSR mutations in Midwestern US	24 (0.16)	71 (0.47)	12 (0.08)	0	0	1 (0.0067)	0	0	0

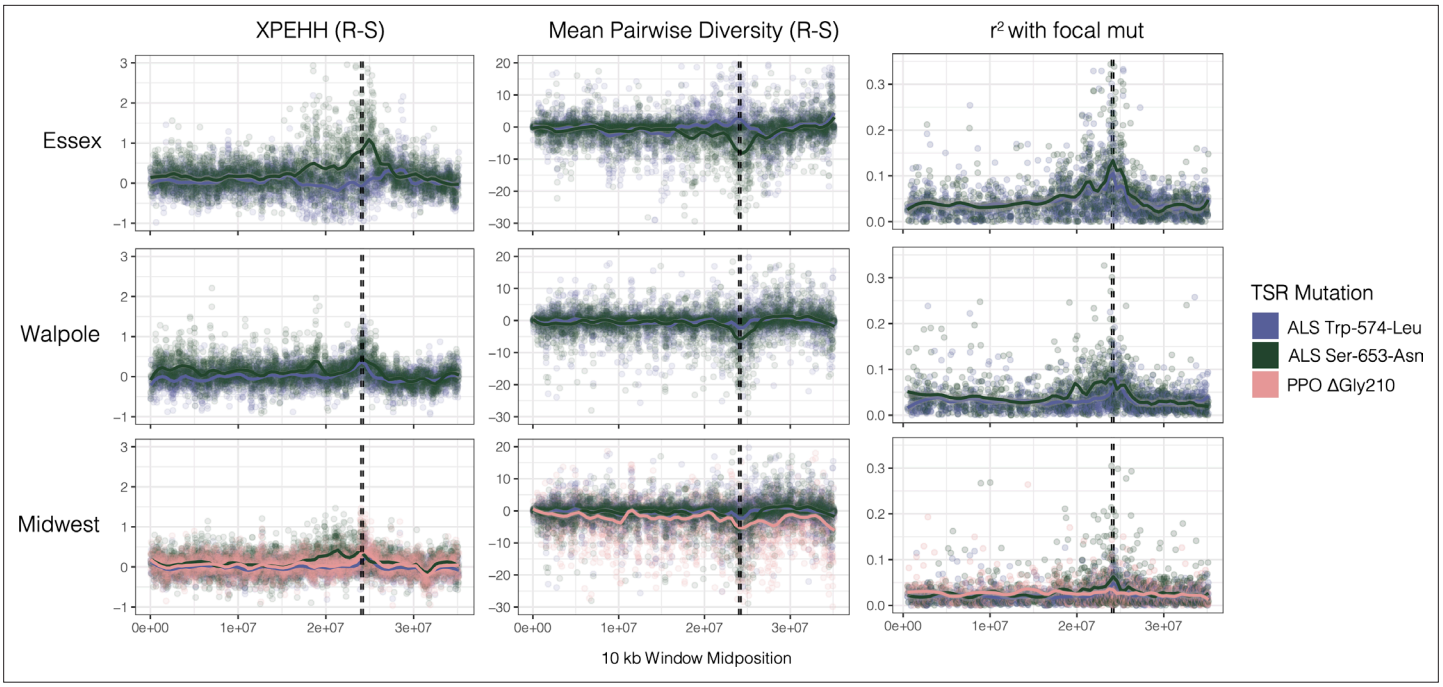

Figure 1. Sweep-scan summary statistics by geographic region. (Left) Difference in integrated haplotype homozygosities (XPEHH) between haplotypes carrying the focal TSR mutation and susceptible haplotypes. (Middle) Difference in mean pairwise diversity between haplotypes carrying the focal TSR mutation and susceptible haplotypes. (Right) r^2 of other missense mutations with focal TSR mutation 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.

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 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 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 harboring 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, the majority of populations (5/8 populations within the Midwestern US, 5/5 populations in Essex County, and 4/6 populations in Walpole) harbor 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 are located only ~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. We found that corresponding selective sweep signals appear to be highly heterogeneous across geographic regions and across resistance mutations (***Figure 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, ***Figure 1*** top-left green line). 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) (***Figure 1***, top middle/right). In contrast, the Trp-574-Leu mutation in Essex actually shows a slight excess of heterozygosity and excess diversity compared to susceptible haplotypes, but nearly as strong LD with other missense SNPs (***Figure 1***, purple line top row).

Selective sweep signals are much subtler in Walpole and especially in the Midwestern US compared to Essex. In Walpole as in Essex, 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 (***Figure 1***, pink line bottom row).

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 (***Foes et al., 2017***; ***Patzoldt and Tranel, 2017***; ***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 resistance 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, 1997***; ***Griffiths and Marjoram, 1996***; ***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 (***Hubisz and Siepel, 2020***; ***Rasmussen et al., 2014***). We assessed the likelihood of the ARG inferences under varying constant recombination rates and over two time step parameters (***Figure 2—figure supplements 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

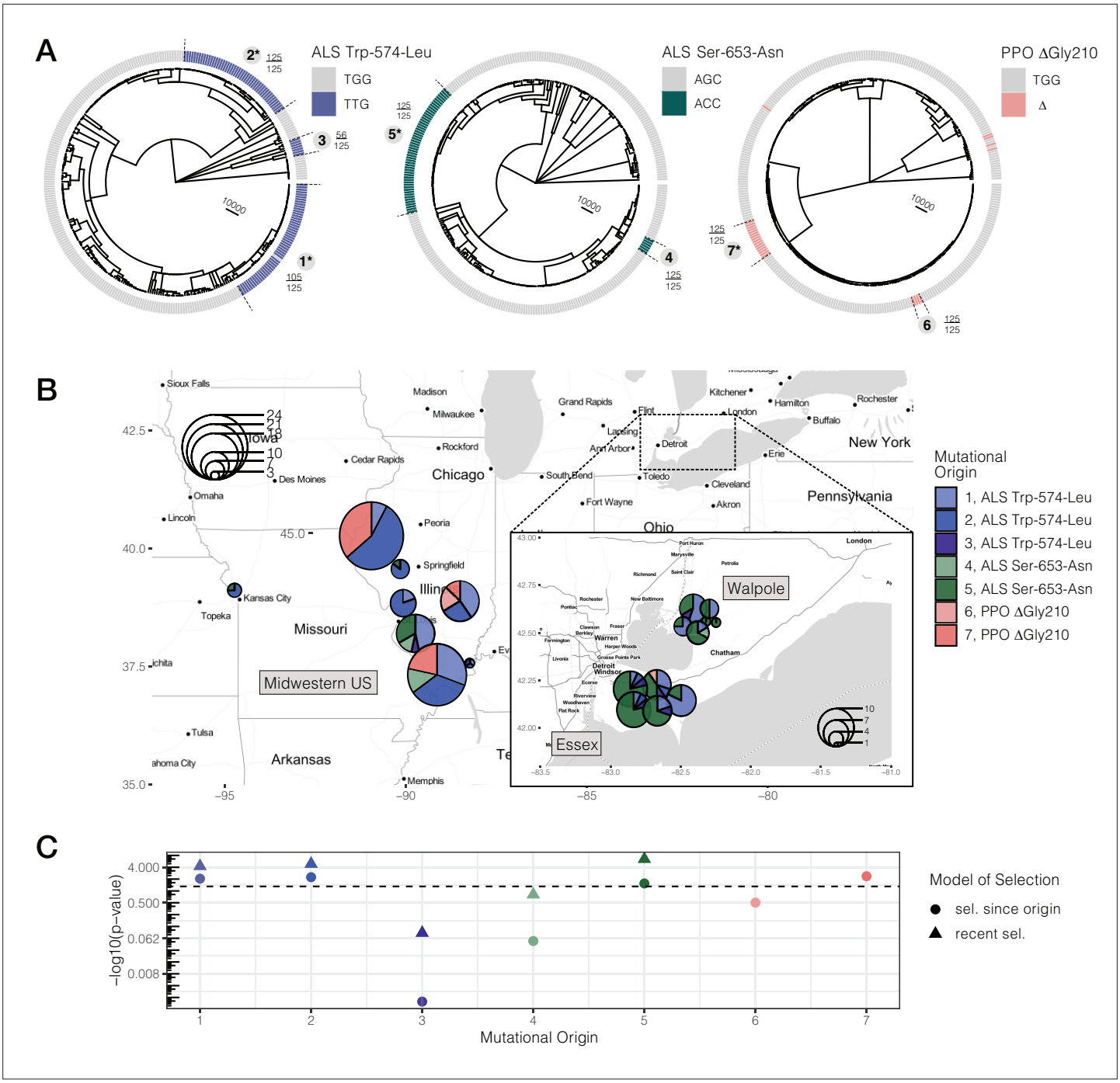

Figure 2. Repeated-independent origins and range-wide distribution of three target-site resistance mutations, along with their associated significance of selection over two different timescales. (**A**) Trees at focal TSR loci corresponding to an ARG estimated across 20 kb SNPs. Bold numbers around trees 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 1250 MCMC iterations is depicted by the fraction found outside each cluster. (**B**) Geographic distribution of haplotypes originating from distinct mutational lineages as inferred from A. 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 $\alpha = 0.05$, after false discovery rate correction.

The online version of this article includes the following source data and figure supplement(s) for figure 2:

Source data 1. Tree-based coalescent test for selection under two scenarios; selection on since the mutation first arose and selection even more recent

Figure 2 continued on next page

Figure 2 continued

timescales (i.e.the last 1% of the tree).

Source data 2. Tree sequence corresponding to ALS Trp-574-Leu, ALS Ser-653-Asn, and PPO ΔGly210 extracted from the most likely iteration of the ARGweaver MCMC.

Source data 3. Resistance status at ALS Trp-574-Leu, ALS Ser-653-Asn, and PPO ΔGly210 for haplotypes mapped in ***Figure 2***.

Figure supplement 1. Bootstrapped gene trees of ALS (3 kb) and PPO (10 kb) (coding sequence + 1 kb on either side) alongside TSR mutations across all 162 individuals.

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 1250 MCMC iterations.

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 supplement 4. Phased haplotypes corresponding to distinct origins of target-site resistance mutations at the ALS Trp-574-Leu (Left side, grey vertical dashed line), ALS Ser-653-Asn (left side, black vertical dashed line) and PPO ΔGly210 (right side, dashed vertical white line) positions, relative to susceptible lineages.

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 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).

mutations, and most specific origin scenarios, are consistent even across less likely recombination rate parameterizations (***Figure 2—figure supplements 2 and 3***). Compared to the gene trees based on the average mutational history across the PPO and ALS (***Figure 2—figure supplements 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 supplements 4–6***).

When haplotypes belonging to distinct mutational lineages are mapped across populations (***Figure 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 exclusive to the Midwestern US (44/47 resistant haplotypes mapping to the Midwest or 94%; 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 and have not yet had the chance to spread extensively.

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 (Materials and 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.

Under the scenario of consistent selection since the origin of the mutation, four out of seven mutational origins we tested were significant at $\alpha = 0.05$ after a 5% false discovery rate (FDR) correction (***Figure 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 ($p = 0.0089$),

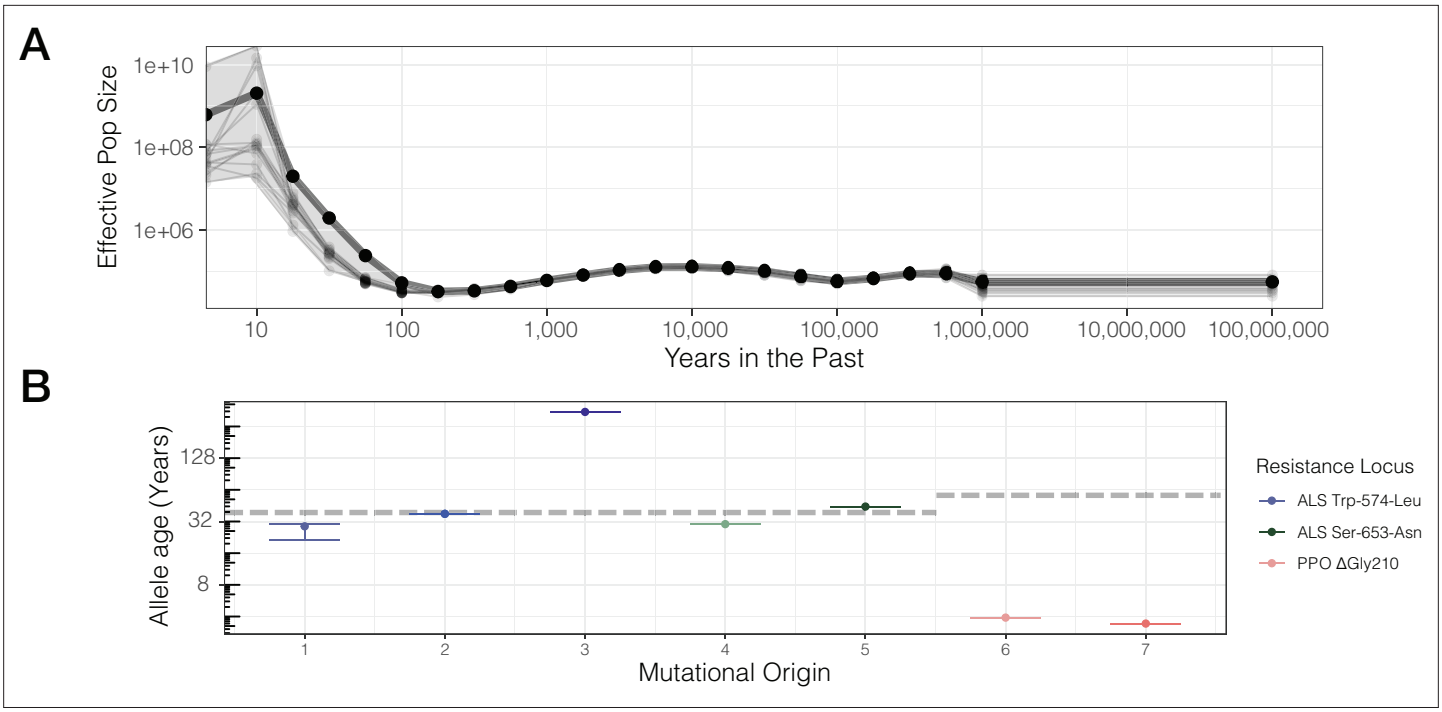

Figure 3. Contemporary population expansion of *A. tuberculatus* and corresponding ages of TSR variants. (**A**) Relate-inferred effective population size through time, illustrating a remarkable population expansion occurring over the last 100 years. The bold line indicates results from genome-wide SNPs, 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.

The online version of this article includes the following source data for figure 3:

Source data 1. Effective population size estimates for each chromosome, and genome-wide, from Relate.

Source data 2. Unscaled allele age estimates corresponding to the seven inferred origins of ALS Trp-574-Leu, ALS Ser-653-Asn, and PPO ΔGly210.

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 1% of the tree showed that for three of the five selected lineages that predated this time cutoff, there is even stronger evidence for selection on more recent timescales, after FDR correction (***Figure 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 × 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 δaδi (***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 ~500,000; recent epoch var. *rudis* N_e ~5,000,000; ***Kreiner et al., 2019***), we now used Relate to infer effective population size through time from genome-wide tree sequence data on even more recent timescales (***Speidel et al., 2021***; ***Speidel et al., 2019***). Historical N_e between 100 and 1,000,000 years ago appears to have stayed relatively consistent, with a harmonic mean of ~63,000 (standard error across chromosomes ± 7000 years)

(***Figure 3A***). However, our samples show evidence for massive recent population expansion over the last 100 years, with the contemporary geometric mean N_e estimate 3–4 orders of magnitude higher than the historical N_e (***Figure 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 1250 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 (***Figure 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 (***Figure 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 their associated signatures of selection. The spread of these parallel origins across the landscape further allows us to observe how these alleles interact when 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 target-site resistance 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 supplements 1 and 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 × 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, suggests 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 (≤ 500 bp) across the genome, this level of repulsion is unexpected (one-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 (***Figure 4A***). This long-range repulsion is not seen for other pairs of loci that are outliers for particularly strong repulsion within 500 bps (***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 (***Figure 4B***, ***Figure 4—figure supplement 3***). Together, this indicates that selection for these alternative resistance variants from ALS-inhibiting 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

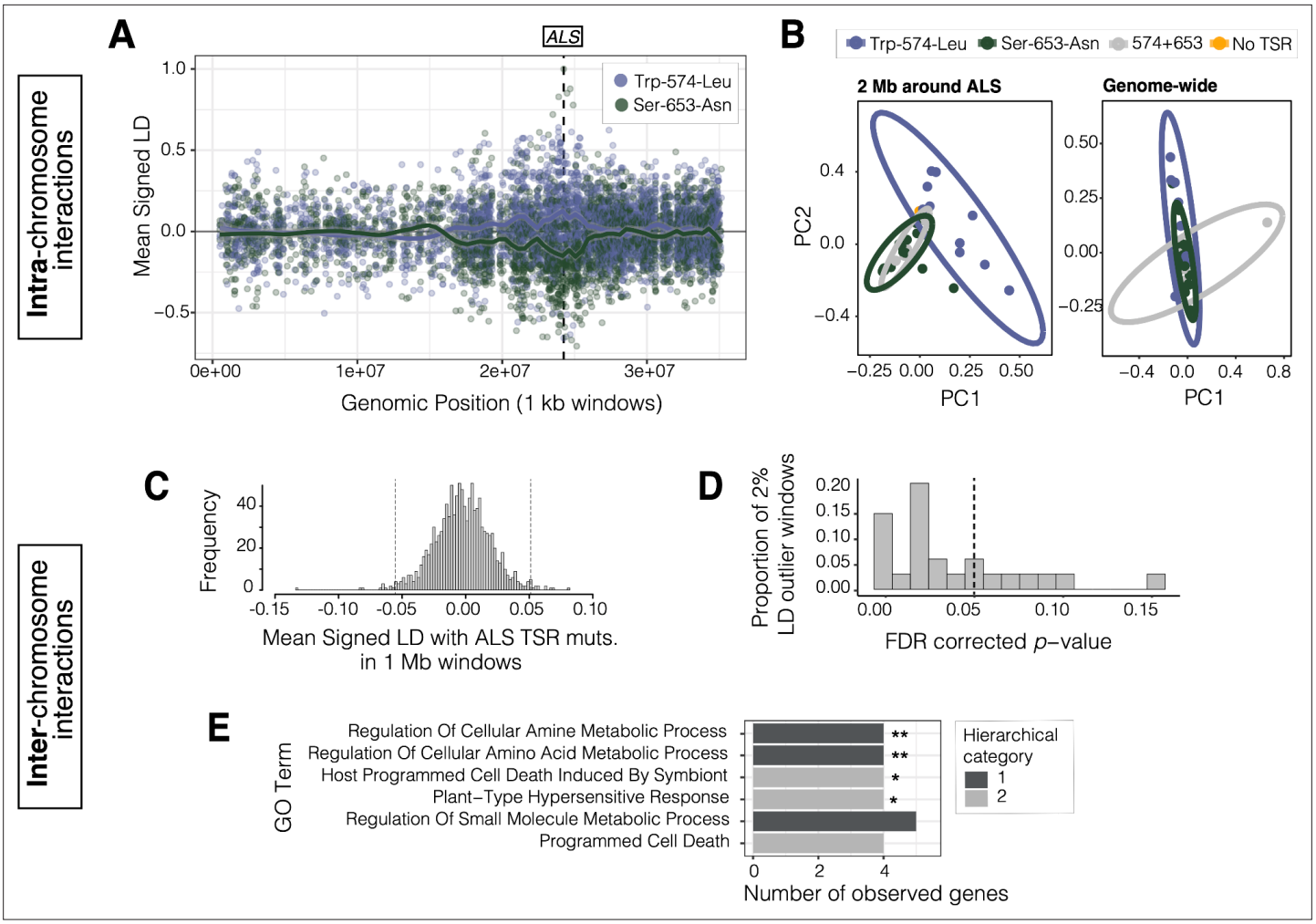

Figure 4. Signals of intra- and inter-chromosomal allelic interactions with target-site resistance mutations. (**A**) The breadth of haplotype competition between TSR mutations, illustrated by repulsion linkage disequilibrium (opposite signed LD, *r*) between two target-site-resistance mutations and 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 574 or 653) with 1 Mb windows genome-wide in Essex, excluding the ALS containing Scaffold 11. 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 within genomic windows and recalculating LD 1,000 times. (**E**) GO terms significantly enriched for biological process after FDR correction from the set of 348 genes mapping to the top 13, 1 Mb windows that show significantly extreme LD with ALS TSR mutations in Essex. Number of asterisks represent significance level after bonferroni correction (** = p < 0.01, * = p < 0.05).

The online version of this article includes the following source data and figure supplement(s) for figure 4:

Source data 1. Signed LD in 1 kb windows with two target-site-resistance mutations and bi-allelic missense SNPs surrounding them on Scaffold 11 in Essex.

Source data 2. Local PCA at 2 Mb around ALS, along with a genome-wide PCA (excluding the ALS containing scaffold 11), in Essex.

Source data 3. Signed LD of 1 Mb windows across the genome (aside from scaffold 11) with either ALS Trp-574-Leu and ALS Ser-653-Asn.

Figure supplement 1. Signed LD *(r)* between pairs of missense (nonsynonymous) and synonymous SNPs of similar frequency (minor allele frequency > 0.20) and physical distance (< 500 bp) as ALS Trp-574-Leu and ALS Ser-653-Asn across the genome.

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.

Figure supplement 3. Signature of population structure for 10 Mb around ALS based on PCAs of genotypes in Essex.

Figure supplement 4. Correlation of signed LD (*r*) between two target-site resistance mutations, and synonymous or missense mutations across the genome.

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; ***Busi et al., 2013***; ***ffrench-Constant et al., 2004***; ***Kreiner et al., 2020***; ***Petit et al., 2010***; ***Preston, 2003***). 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 (***Busi et al., 2013***; ***ffrench-Constant et al., 2004***; ***Kreiner et al., 2020***; ***Petit et al., 2010***; ***Preston, 2003***) 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 (***Figure 4C***). We then focused on the upper 1% and lower 1% of windows with particularly extreme signed LD with either TSR mutation (24/1156 1 Mb windows). On each outlier window, we performed a permutation in which we randomized TSR allele assignment among individuals 1000 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 (one-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 $\alpha = 0.05$ (***Figure 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 (***Figure 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 signaling (***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, 2014***). 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 three common 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 (***Figure 2—figure supplement 1***), we show that accounting for intragenic recombination through ARG inference has been 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* $4N_e\mu \sim \Theta_\pi = 0.041$ 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 $\Theta_\pi/2$—once every six generations (i.e. 0.041/2 x 8 known TSR loci = 0.164 TSR mutations per generation; see also ***Charlesworth, 2009***; ***Karasov et al., 2010***). 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 N_e 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 N_e over the last 50 years (~8×10^8) 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 $\Theta \sim 1$ 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 (***Figure 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 (***Figure 3A***)—population size increasing by four orders of magnitude over the last 100 years. This expansion also corresponds well with *A. tuberculatus*'s contemporary agricultural association (***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 modeling 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. Nonetheless, 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 (***Figure 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 N_e 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 N_e 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 (***Salas et al., 2016***; ***Tranel and Wright, 2002***). 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 (***Patzoldt and Tranel, 2017***). 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 both physically linked and unlinked loci, 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 (***Figure 4A and 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 $\rho = 4\ N_e r = 0.0575$ 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 bp x (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 interference (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 (***Figure 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, ***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. ***Heap, 2014***; ***Menchari et al., 2006***) (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 neighboring 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.

Materials and 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 eight 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., 2017***). 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., 2020**).

SNP calling and phasing genotypes

Filtered VCFs were obtained from **Kreiner et al., 2019** for all analyses. Briefly, freebayes-called SNPs were filtered based on missing data (> 80% present), repeat content, allelic bias (> 0.25 and < 0.75), read paired status, and mapping quality (> Q30). Six individuals were removed due to excess missing data, leaving 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 eight ALS codons (**Yu and Powles, 2014**), three PPO codons (**Giacomini et al., 2017**; **Rousonelos et al., 2017**; **Varanasi et al., 2017**), one PsbA (conferring resistance to photosystem II inhibitors) codon (**Lu et al., 2019**), and three EPSPS codons (**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 ($100/4N_e r$; 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 modeling in dadi (**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 inference of 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 (**Rédei, 2008**) 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) at each tip of both gene trees using ggtree (**Yu, 2020**).

We ran ARGweaver (**Hubisz and Siepel, 2020**; **Rasmussen et al., 2014**) 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) 30 `--maxtime` (max coalescent time) $100e^{3}$ `-c` (bp compression rate) 1 `-n` (MCMC samples) 1,250. We used an effective population size of 500,000, based on the best fitting demographic model previously inferred from this dataset with δaδi (**Gutenkunst et al., 2010**; **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 $r = 7e^{-7}$ to $7e^{-9}$, drawing our inferences from the parameter values that maximized the likelihood of observing our data ($r = 7e^{-8}$). Similarly, we also tested out two parameter values of --ntimes (20, 30), and used t = 30 for our inference, together with $r = 7e^{-8}$. 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 an SNP,

changing the reference allele at this site to a single bp record matching the reference call at the start position (G) with 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. To do so, 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., 2021***; ***Speidel et al., 2019***), 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 probability that a mutation spreads to at least a given frequency, from f_N to N-k_s + 2.

$$p_{R,de\,novo} = \sum_{f=f_N}^{N-k_S+2} \frac{(f-1)\left(\frac{N-f-1}{k_S-3}\right)}{\frac{N-1}{k_S-1}} \tag{1}$$

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 $k_S(t)$.

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

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 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), reestimate branch lengths with this updated mutation rate (ReEstimateBranchLengths), and then lastly estimate 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 and 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. 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 Petrov, 2016***; ***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 with all missense alleles for each of 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 et al., 2012***), for genotypes spanning 2 Mb around ALS, 10 Mb around ALS, and 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 (< 500 bp). 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 1000 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, and our two reviewers, Pleuni Pennings and Philipp Messer. 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. JMK was supported

by the Rosemary Grant Advanced Award from the Society for the Study of Evolutionary Biology and NSERC PGS-D, DW was supported by the Max Planck Society, JRS and SIW were supported by Discovery Grants from NSERC Canada, and SIW was additionally supported by a Canada Research Chair in Population Genomics.

Additional information

Competing interests

Detlef Weigel: Deputy editor, eLife. The other authors declare that no competing interests exist.

Funding

Funder	Grant reference number	Author
Natural Sciences and Engineering Research Council of Canada	PGS-D	Julia M Kreiner
Society for the Study of Evolution	Rosemary Grant Advanced Award	Julia M Kreiner
Natural Sciences and Engineering Research Council of Canada	Discovery Grant	John Stinchcombe Stephen I Wright
Canada Research Chairs	Population Genomics	Stephen I Wright
Max Planck Society		Detlef Weigel

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Julia M Kreiner, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Visualization, Writing – original draft, Writing – review and editing; George Sandler, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review and editing; Aaron J Stern, Methodology, Software; Patrick J Tranel, Detlef Weigel, Data curation, Funding acquisition, Writing – review and editing; John R Stinchcombe, Data curation, Funding acquisition, Methodology, Supervision, Writing – original draft, Writing – review and editing; Stephen I Wright, Conceptualization, Data curation, Funding acquisition, Methodology, Supervision, Writing – original draft, Writing – review and editing

Author ORCIDs

Julia M Kreiner http://orcid.org/0000-0002-8593-1394
George Sandler http://orcid.org/0000-0001-9420-1521
Aaron J Stern http://orcid.org/0000-0001-7368-5520
Detlef Weigel http://orcid.org/0000-0002-2114-7963
John R Stinchcombe http://orcid.org/0000-0003-3349-2964
Stephen I Wright http://orcid.org/0000-0001-9973-9697

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.70242.sa1
Author response https://doi.org/10.7554/eLife.70242.sa2

Additional files

Supplementary files

- Transparent reporting form

Data availability

Sequencing data used in this paper were previously deposited in ENA under project number PRJEB31711, and reference genome is available on CoGe (reference ID = 54057). Code used to generate results in this paper is available at https://github.com/jkreinz/TSRevolution (copy archived at swh:1:rev:93ccf9e9d8d1fa25d1e4ef37a62bbb3c91df35dd).

The following previously published datasets were used:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Kreiner JM, Giacomini D, Bemm F, Waithaka B, Regalado J, Lanz C, Hildebrandt J, Sikkema PH, Tranel PJ, Weigel D, Stinchcombe JR, Wright SI	2019	Multiple modes of convergent adaptation in the spread of glyphosate-resistant Amaranthus tuberculatus	https://www.ebi.ac.uk/ena/browser/view/PRJEB31711	EBI, PRJEB31711

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**:403–410. DOI: https://doi.org/10.1016/S0022-2836(05)80360-2, PMID: 2231712

Baena-González E, Rolland F, Thevelein JM, Sheen J. 2007. A central integrator of transcription networks in plant stress and energy signalling. *Nature* **448**:938–942. DOI: https://doi.org/10.1038/nature06069, PMID: 17671505

Baker HG. 1974. The Evolution of Weeds. *Annual Review of Ecology and Systematics* **5**:1–24. DOI: https://doi.org/10.1146/annurev.es.05.110174.000245

Baucom RS. 2019. Evolutionary and ecological insights from herbicide-resistant weeds: what have we learned about plant adaptation, and what is left to uncover? *New Phytologist* **223**:68–82. DOI: https://doi.org/10.1111/nph.15723, PMID: 30710343

Bell G. 2013. Evolutionary rescue and the limits of adaptation. *Philosophical Transactions of the Royal Society B* **368**:20120080. DOI: https://doi.org/10.1098/rstb.2012.0080, PMID: 23209162

Brown HM. 1990. Mode of action, crop selectivity, and soil relations of the sulfonylurea herbicides. *Pesticide Science* **29**:263–281. DOI: https://doi.org/10.1002/ps.2780290304

Busi R, Neve P, Powles S. 2013. Evolved polygenic herbicide resistance in Lolium rigidum by low-dose herbicide selection within standing genetic variation. *Evolutionary Applications* **6**:231–242. DOI: https://doi.org/10.1111/j.1752-4571.2012.00282.x, PMID: 23798973

Charlesworth B. 2009. Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics* **10**:195–205. DOI: https://doi.org/10.1038/nrg2526, PMID: 19204717

Chatham LA, Wu C, Riggins CW, Hager AG, Young BG, Roskamp GK, Tranel PJ. 2017. EPSPS Gene Amplification is Present in the Majority of Glyphosate-Resistant Illinois Waterhemp (Amaranthus tuberculatus) Populations. *Weed Technology* **29**:48–55. DOI: https://doi.org/10.1614/WT-D-14-00064.1

Chen L, Su ZZ, Huang L, Xia FN, Qi H, Xie LJ, Xiao S, Chen QF. 2017. The AMP-Activated Protein Kinase KIN10 Is Involved in the Regulation of Autophagy in Arabidopsis. *Frontiers in Plant Science* **8**:1201. DOI: https://doi.org/10.3389/fpls.2017.01201, PMID: 28740502

Comai L, Stalker D. 1986. Mechanism of action of herbicides and their molecular manipulation. *Surveys of Plant Molecular and Cell Biology* **3**:166–195.

Cooper TF. 2007. Recombination speeds adaptation by reducing competition between beneficial mutations in populations of *Escherichia coli*. *PLOS Biology* **5**:e225. DOI: https://doi.org/10.1371/journal.pbio.0050225

Costea M, Weaver SE, Tardif FJ. 2005. The Biology of Invasive Alien Plants in Canada. 3. *Amaranthus tuberculatus* (Moq.) Sauer var. *rudis* (Sauer) Costea & Tardif. *Canadian Journal of Plant Science* **85**:507–522. DOI: https://doi.org/10.4141/P04-101

Dayan FE, Barker A, Tranel PJ. 2018. Origins and structure of chloroplastic and mitochondrial plant protoporphyrinogen oxidases: implications for the evolution of herbicide resistance. *Pest Management Science* **74**:2226–2234. DOI: https://doi.org/10.1002/ps.4744

Delaneau O, Zagury JF, Marchini J. 2013. Improved whole-chromosome phasing for disease and population genetic studies. *Nature Methods* **10**:5–6. DOI: https://doi.org/10.1038/nmeth.2307, PMID: 23269371

Delaneau O, Zagury JF, Robinson MR, Marchini JL, Dermitzakis ET. 2019. Accurate, scalable and integrative haplotype estimation. *Nature Communications* **10**:1–10. DOI: https://doi.org/10.1038/s41467-019-13225-y, PMID: 31780650

Délye C. 2013. Unravelling the genetic bases of non-target-site-based resistance (NTSR) to herbicides: a major challenge for weed science in the forthcoming decade: Unravelling the genetic bases of non-target-site-based resistance to herbicides. *Pest Management Science* **69**:176–187. DOI: https://doi.org/10.1002/ps.3318

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* **16**:157. DOI: https://doi.org/10.1186/s13059-015-0721-2, PMID: 26243257

Faus I, Zabalza A, Santiago J, Nebauer SG, Royuela M, Serrano R, Gadea J. 2015. Protein kinase GCN2 mediates responses to glyphosate in Arabidopsis. *BMC Plant Biology* **15**:14. DOI: https://doi.org/10.1186/s12870-014-0378-0, PMID: 25603772

Feder AF, Rhee SY, Holmes SP, Shafer RW, Petrov DA, Pennings PS. 2016. More effective drugs lead to harder selective sweeps in the evolution of drug resistance in HIV-1. *eLife* **5**:e10670. DOI: https://doi.org/10.7554/eLife.10670

Feder AF, Kline C, Polacino P, Cottrell M, Kashuba ADM, Keele BF, Hu SL, Petrov DA, Pennings PS, Ambrose Z. 2017. A spatio-temporal assessment of simian/human immunodeficiency virus (SHIV) evolution reveals a highly dynamic process within the host. *PLOS Pathogens* **13**:e1006358. DOI: https://doi.org/10.1371/journal.ppat.1006358

Feder AF, Pennings PS, Hermisson J, Petrov DA. 2019. Evolutionary Dynamics in Structured Populations Under Strong Population Genetic Forces. *G3: Genes, Genomes, Genetics* **9**:3395–3407. DOI: https://doi.org/10.1534/g3.119.400605

ffrench-Constant RH, Daborn PJ PJ, Goff GL. 2004. The genetics and genomics of insecticide resistance. *Trends in Genetics: TIG* **20**:163–170. DOI: https://doi.org/10.1016/j.tig.2004.01.003

Flood PJ, Heerwaarden J, Becker F, Snoo CB, Harbinson J, Aarts MGM. 2016. Whole-Genome Hitchhiking on an Organelle Mutation. *Current Biology* **26**:1306–1311. DOI: https://doi.org/10.1016/j.cub.2016.03.027

Foes MJ, Liu L, Tranel PJ, Wax LM, Stoller EW. 2017. A biotype of common waterhemp (Amaranthus rudis) resistant to triazine and ALS herbicides. *Weed Science* **46**:514–520. DOI: https://doi.org/10.1017/S0043174500091013

Fujiki Y, Yoshikawa Y, Sato T, Inada N, Ito M, Nishida I, Watanabe A. 2001. Dark-inducible genes from *Arabidopsis thaliana* are associated with leaf senescence and repressed by sugars. *Physiologia Plantarum* **111**:345–352. DOI: https://doi.org/10.1034/j.1399-3054.2001.1110312.x, PMID: 11240919

Garud NR, Rosenberg NA. 2015. Enhancing the mathematical properties of new haplotype homozygosity statistics for the detection of selective sweeps. *Theoretical Population Biology* **102**:94–101. DOI: https://doi.org/10.1016/j.tpb.2015.04.001, PMID: 25891325

Garud NR, Petrov DA. 2016. Elevated Linkage Disequilibrium and Signatures of Soft Sweeps Are Common in *Drosophila melanogaster*. *Genetics* **203**:863–880. DOI: https://doi.org/10.1534/genetics.115.184002, PMID: 27098909

Giacomini DA, Umphres AM, Nie H, Mueller TC, Steckel LE, Young BG, Scott RC, Tranel PJ. 2017. Two new PPX2 mutations associated with resistance to PPO-inhibiting herbicides in Amaranthus palmeri. *Pest Management Science* **73**:1559–1563. DOI: https://doi.org/10.1002/ps.4581, PMID: 28370968

Good BH. 2020. Linkage Disequilibrium between Rare Mutations. Cold Spring Harbor Laboratory. DOI: https://doi.org/10.1101/2020.12.10.420042

Green JM. 2007. Review of Glyphosate and Als-inhibiting Herbicide Crop Resistance and Resistant Weed Management. *Weed Technology* **21**:547–558. DOI: https://doi.org/10.1614/WT-06-004.1

Griffiths RC, Marjoram P. 1996. Ancestral inference from samples of DNA sequences with recombination. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* **3**:479–502. DOI: https://doi.org/10.1089/cmb.1996.3.479

Griffiths RC, Marjoram P. 1997. An Ancestral Recombination Graph. Tavare S (Ed). *Progress in Population Genetics and Human Evolution*. Springer. p. 16–329. DOI: https://doi.org/10.1007/978-1-4757-2609-1_16

Griffiths RC, Tavaré S. 1998. The age of a mutation in a general coalescent tree. *Communications in Statistics. Stochastic Models* **14**:273–295. DOI: https://doi.org/10.1080/15326349808807471

Gupta S, Harkess A, Soble A, Van Etten M, Leebens-Mack J, Baucom RS. 2021. Inter-Chromosomal Linkage Disequilibrium and Linked Fitness Cost Loci Influence the Evolution of Nontarget Site Herbicide Resistance in an Agricultural Weed. *bioRxiv*. DOI: https://doi.org/10.1101/2021.04.04.438381

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLOS Genetics* **5**:e1000695. DOI: https://doi.org/10.1371/journal.pgen.1000695, PMID: 19851460

Gutenkunst R, Hernandez R, Williamson S, Bustamante C. 2010. Diffusion Approximations for Demographic Inference: DaDi. *Nature Precedings* **1**:4594. DOI: https://doi.org/10.1038/npre.2010.4594.1

Harris RS. 2007. Improved pairwise alignmnet of genomic DNA. Computer Science and Engineering. http://www.bx.psu.edu/~rsharris/rsharris_phd_thesis_2007.pdf

Heap I. 2014. Herbicide Resistant Weeds. Pimentel D, Peshin R (Eds). *Integrated Pest Management*. Springer Netherlands. p. 281–301. DOI: https://doi.org/10.1007/978-94-007-7796-5_12

Heap I. 2021. The international herbicide-resistant weed database. http://www.weedscience.org/Home.aspx [Accessed January 25, 2021].

Hejase HA, Salman-Minkov A, Campagna L, Hubisz MJ, Lovette IJ, Gronau I, Siepel A. 2020. Genomic islands of differentiation in a rapid avian radiation have been driven by recent selective sweeps. *PNAS* **117**:30554–30565. DOI: https://doi.org/10.1073/pnas.2015987117, PMID: 33199636

Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. *Genetical Research* **8**:269–294. DOI: https://doi.org/10.1017/s0016672300010156, PMID: 5980116

Holst N, Rasmussen IA, Bastiaans L. 2007. Field weed population dynamics: a review of model approaches and applications. *Weed Research* **47**:1–14. DOI: https://doi.org/10.1111/j.1365-3180.2007.00534.x

Hubisz M, Siepel A. 2020. Inference of Ancestral Recombination Graphs Using ARGweaver. *Methods in Molecular Biology* **2090**:231–266. DOI: https://doi.org/10.1007/978-1-0716-0199-0_10, PMID: 31975170
Hudson RR. 1983. Properties of a neutral allele model with intragenic recombination. *Theoretical Population Biology* **23**:183–201. DOI: https://doi.org/10.1016/0040-5809(83)90013-8, PMID: 6612631
Karasov T, Messer PW, Petrov DA, Malik HS. 2010. Evidence that Adaptation in *Drosophila* Is Not Limited by Mutation at Single Sites. *PLOS Genetics* **6**:e1000924. DOI: https://doi.org/10.1371/journal.pgen.1000924
Kent TV, Uzunović J, Wright SI. 2017. Coevolution between transposable elements and recombination. *Philosophical Transactions of the Royal Society B* **372**:20160458. DOI: https://doi.org/10.1098/rstb.2016.0458
Kreiner JM, Stinchcombe JR, Wright SI. 2018. Population Genomics of Herbicide Resistance: Adaptation via Evolutionary Rescue. *Annual Review of Plant Biology* **69**:611–635. DOI: https://doi.org/10.1146/annurev-arplant-042817-040038, PMID: 29140727
Kreiner JM, Giacomini DA, Bemm F, Waithaka B, Regalado J, Lanz C, Hildebrandt J, Sikkema PH, Tranel PJ, Weigel D, Stinchcombe JR, Wright SI. 2019. Multiple modes of convergent adaptation in the spread of glyphosate-resistant Amaranthus tuberculatus. *PNAS* **116**:21076–21084. DOI: https://doi.org/10.1073/pnas.1900870116, PMID: 31570613
Kreiner JM, Tranel PJ, Weigel D, Stinchcombe JR. 2020. The Genetic Architecture and Genomic Context of Glyphosate Resistance. *bioRxiv*. DOI: https://doi.org/10.1101/2020.08.19.257972
Kreiner JM, Tranel PJ, Weigel D, Stinchcombe JR, Wright SI. 2021. The genetic architecture and population genomic signatures of glyphosate resistance in Amaranthus tuberculatus. *Molecular Ecology* **30**:5373–5389. DOI: https://doi.org/10.1111/mec.15920, PMID: 33853196
Kuester A, Fall E, Chang SM, Baucom RS. 2017. Shifts in outcrossing rates and changes to floral traits are associated with the evolution of herbicide resistance in the common morning glory. *Ecology Letters* **20**:41–49. DOI: https://doi.org/10.1111/ele.12703, PMID: 27905176
Küpper A, Manmathan HK, Giacomini D, Patterson EL, McCloskey WB, Gaines TA. 2018. Population Genetic Structure in Glyphosate-Resistant and -Susceptible Palmer Amaranth (Amaranthus palmeri) Populations Using Genotyping-by-sequencing (GBS). *Frontiers in Plant Science* **9**:29. DOI: https://doi.org/10.3389/fpls.2018.00029, PMID: 29422910
Lenormand T, Harmand N, Gallet R. 2018. Cost of resistance: an unreasonably expensive concept. *Rethinking Ecology* **3**:51–70. DOI: https://doi.org/10.3897/rethinkingecology.3.31992
Leslie T, Baucom RS. 2014. De Novo Assembly and Annotation of the Transcriptome of the Agricultural Weed Ipomoea purpurea Uncovers Gene Expression Changes Associated with Herbicide Resistance. *G3: Genes, Genomes, Genetics* **4**:2035–2047. DOI: https://doi.org/10.1534/g3.114.013508
Liu J, Davis AS, Tranel PJ. 2017. Pollen Biology and Dispersal Dynamics in Waterhemp (Amaranthus *tuberculatus*) *Weed Science* **60**:416–422. DOI: https://doi.org/10.1614/WS-D-11-00201.1
Lu H, Yu Q, Han H, Owen MJ, Powles SB. 2019. A novel psbA mutation (Phe274-Val) confers resistance to PSII herbicides in wild radish (Raphanus raphanistrum). *Pest Management Science* **75**:144–151. DOI: https://doi.org/10.1002/ps.5079
MacLean RC, Hall AR, Perron GG, Buckling A. 2010. The population genetics of antibiotic resistance: integrating molecular mechanisms and treatment contexts. *Nature Reviews. Genetics* **11**:405–414. DOI: https://doi.org/10.1038/nrg2778, PMID: 20479772
Madeira F, Park Y, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* **47**:W636–W641. DOI: https://doi.org/10.1093/nar/gkz268
Martin M, Patterson M, Garg S, Fischer SO, Pisanti N, Klau GW, Schöenhuth A, Marschall T. 2016. WhatsHap: Fast and Accurate Read-Based Phasing. *bioRxiv*. DOI: https://doi.org/10.1101/085050
Mather K. 1969. Selection through competition. *Heredity* **24**:529–540. DOI: https://doi.org/10.1038/hdy.1969.76
Mazur BJ, Falco SC. 1989. The Development of Herbicide Resistant Crops. *Annual Review of Plant Physiology and Plant Molecular Biology* **40**:441–470. DOI: https://doi.org/10.1146/annurev.pp.40.060189.002301
McNaughton KE, Lee EA, Tardif FJ. 2001. Mutations in the ALS Gene Conferring Resistance to Group II Herbicides in Redroot Pigweed (Amaranthus Retroflexus) and Green Pigweed (A Powellii). *Weed Science* **53**:17–22. DOI: https://doi.org/10.1614/WS-04-109
Menchari Y, Camilleri C, Michel S, Brunel D, Dessaint F, Le Corre V, Délye C. 2006. Weed response to herbicides: regional-scale distribution of herbicide resistance alleles in the grass weed Alopecurus myosuroides. *The New Phytologist* **171**:861–873. DOI: https://doi.org/10.1111/j.1469-8137.2006.01788.x, PMID: 16918556
Molin WT, Wright AA, Lawton-Rauh A, Saski CA. 2017. The unique genomic landscape surrounding the EPSPS gene in glyphosate resistant Amaranthus palmeri: a repetitive path to resistance. *BMC Genomics* **18**:91. DOI: https://doi.org/10.1186/s12864-016-3336-4, PMID: 28095770
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. 2020. Draft Genomes of Amaranthus tuberculatus, Amaranthus hybridus, and Amaranthus palmeri. *Genome Biology and Evolution* **12**:1988–1993. DOI: https://doi.org/10.1093/gbe/evaa177, PMID: 32835372
Murphy BP, Tranel PJ. 2019. Target-Site Mutations Conferring Herbicide Resistance. *Plants* **8**:E382. DOI: https://doi.org/10.3390/plants8100382, PMID: 31569336
Nakka S, Thompson CR, Peterson DE, Jugulam M. 2017. Target Site–Based and Non–Target Site Based Resistance to ALS Inhibitors in Palmer Amaranth (Amaranthus palmeri). *Weed Science* **65**:681–689. DOI: https://doi.org/10.1017/wsc.2017.43

Naylor REL. 2003. Weed Biology: Encyclopedia of Applied Plant Sciences. University of California. DOI: https://doi.org/10.1016/b0-12-227050-9/00158-7
Neve P, Vila-Aiub M, Roux F. 2009. Evolutionary-thinking in agricultural weed management. *The New Phytologist* **184**:783–793. DOI: https://doi.org/10.1111/j.1469-8137.2009.03034.x, PMID: 19780985
Neve P, Busi R, Renton M, Vila-Aiub MM. 2014. Expanding the eco-evolutionary context of herbicide resistance research. *Pest Management Science* **70**:1385–1393. DOI: https://doi.org/10.1002/ps.3757, PMID: 24723489
Orcaray L, Igal M, Zabalza A, Royuela M. 2011. Role of exogenously supplied ferulic and p-coumaric acids in mimicking the mode of action of acetolactate synthase inhibiting herbicides. *Journal of Agricultural and Food Chemistry* **59**:10162–10168. DOI: https://doi.org/10.1021/jf2025538, PMID: 21870840
Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, Shaw RG, Weigel D, Lynch M. 2010. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* **327**:92–94. DOI: https://doi.org/10.1126/science.1180677, PMID: 20044577
Otto SP. 2021. Selective Interference and the Evolution of Sex. *The Journal of Heredity* **112**:9–18. DOI: https://doi.org/10.1093/jhered/esaa026, PMID: 33047117
Patzoldt WL, Tranel PJ. 2017. Multiple ALS Mutations Confer Herbicide Resistance in Waterhemp (Amaranthus tuberculatus). *Weed Science* **55**:421–428. DOI: https://doi.org/10.1614/WS-06-213.1
Pennings PS, Kryazhimskiy S, Wakeley J. 2014. Loss and recovery of genetic diversity in adapting populations of HIV. *PLOS Genetics* **10**:e1004000. DOI: https://doi.org/10.1371/journal.pgen.1004000, PMID: 24465214
Perotti VE, Larran AS, Palmieri VE, Martinatto AK, Alvarez CE, Tuesca D, Permingeat HR. 2019. A novel triple amino acid substitution in the EPSPS found in a high-level glyphosate-resistant Amaranthus hybridus population from Argentina. *Pest Management Science* **75**:1242–1251. DOI: https://doi.org/10.1002/ps.5303, PMID: 30556254
Peterson MA, Collavo A, Ovejero R, Shivrain V, Walsh MJ. 2018. The challenge of herbicide resistance around the world: a current summary. *Pest Management Science* **74**:2246–2259. DOI: https://doi.org/10.1002/ps.4821, PMID: 29222931
Petit C, Duhieu B, Boucansaud K, Délye C. 2010. Complex genetic control of non-target-site-based resistance to herbicides inhibiting acetyl-coenzyme A carboxylase and acetolactate-synthase in Alopecurus myosuroides Huds. *Plant Science* **178**:501–509. DOI: https://doi.org/10.1016/j.plantsci.2010.03.007
Porse A, Jahn LJ, Ellabaan MMH, Sommer MOA. 2020. Dominant resistance and negative epistasis can limit the co-selection of de novo resistance mutations and antibiotic resistance genes. *Nature Communications* **11**:1199. DOI: https://doi.org/10.1038/s41467-020-15080-8, PMID: 32139686
Powles SB, Yu Q. 2010. Evolution in action: plants resistant to herbicides. *Annual Review of Plant Biology* **61**:317–347. DOI: https://doi.org/10.1146/annurev-arplant-042809-112119, PMID: 20192743
Preston C. 2003. Inheritance and linkage of metabolism-based herbicide cross-resistance in rigid ryegrass (Lolium rigidum). *Weed Science* **51**:4–12. DOI: https://doi.org/10.1614/0043-1745(2003)051[0004:IALOMB]2.0.CO;2
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**:559–575. DOI: https://doi.org/10.1086/519795, PMID: 17701901
Ralph P, Coop G. 2010. Parallel adaptation: one or many waves of advance of an advantageous allele? *Genetics* **186**:647–668. DOI: https://doi.org/10.1534/genetics.110.119594, PMID: 20660645
Rasmussen MD, Hubisz MJ, Gronau I, Siepel A. 2014. Genome-wide inference of ancestral recombination graphs. *PLOS Genetics* **10**:e1004342. DOI: https://doi.org/10.1371/journal.pgen.1004342, PMID: 24831947
Rédei GP. 2008. CLUSTAL W (Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice)Encyclopedia of Genetics, Genomics, Proteomics and Informatics. Dordrecht: Springer Netherlands. DOI: https://doi.org/10.1007/978-1-4020-6754-9_3188
Rousonelos SL, Lee RM, Moreira MS, VanGessel MJ, Tranel PJ. 2017. Characterization of a Common Ragweed (Ambrosia artemisiifolia) Population Resistant to ALS- and PPO-Inhibiting Herbicides. *Weed Science* **60**:335–344. DOI: https://doi.org/10.1614/WS-D-11-00152.1
Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, Schaffner SF, Lander ES, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, et al. 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**:913–918. DOI: https://doi.org/10.1038/nature06250
Salas RA, Burgos NR, Tranel PJ, Singh S, Glasgow L, Scott RC, Nichols RL. 2016. Resistance to PPO -inhibiting herbicide in Palmer amaranth from Arkansas. *Pest Management Science* **72**:864–869. DOI: https://doi.org/10.1002/ps.4241
Sandler G, Wright SI, Agrawal AF. 2020. Using Patterns of Signed Linkage Disequilibria to Test for Epistasis in Flies and Plants. Cold Spring Harbor Laboratory. DOI: https://doi.org/10.1101/2020.11.25.399030
Sauer J. 1957. RECENT MIGRATION AND EVOLUTION OF THE DIOECIOUS AMARANTHS. *Evolution; International Journal of Organic Evolution* **11**:11–31. DOI: https://doi.org/10.1111/j.1558-5646.1957.tb02872.x
Shoup DE, Al-Khatib K, Peterson DE. 2003. Common waterhemp (Amaranthus rudis) resistance to protoporphyrinogen oxidase-inhibiting herbicides. *Weed Science* **51**:145–150. DOI: https://doi.org/10.1614/0043-1745(2003)051[0145:CWARRT]2.0.CO;2
Singh S, Singh V, Salas-Perez RA, Bagavathiannan MV, Lawton-Rauh A, Roma-Burgos N. 2018. Target-site mutation accumulation among ALS inhibitor-resistant Palmer amaranth. *Pest Management Science* **8**:350. DOI: https://doi.org/10.1002/ps.5232

Speidel L, Forest M, Shi S. 2019. A method for genome-wide genealogy estimation for thousands of samples. *Nature Genetics* **51**:1321–1329. DOI: https://doi.org/10.1038/s41588-019-0484-x
Speidel L, Cassidy L, Davies RW, Hellenthal G, Skoglund P. 2021. Inferring Population Histories for Ancient Genomes Using Genome-Wide Genealogies. Cold Spring Harbor Laboratory. DOI: https://doi.org/10.1101/2021.02.17.431573
Szpiech ZA, Hernandez RD. 2014. selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Molecular Biology and Evolution* **31**:2824–2827. DOI: https://doi.org/10.1093/molbev/msu211, PMID: 25015648
Tranel PJ, Wright TR. 2002. Resistance of weeds to ALS-inhibiting herbicides: what have we learned? *Weed Science* **50**:700–712. DOI: https://doi.org/10.1614/0043-1745(2002)050[0700:RROWTA]2.0.CO;2
Tranel PJ, Trucco F. 2009. 21st-century weed science: a call for Amaranthus genomics. Weedy and invasive plant genomics. 53–81. DOI: https://doi.org/10.1002/9780813806198
Tranel PJ. 2021. Herbicide resistance in Amaranthus tuberculatus. *Pest Management Science* **77**:43–54. DOI: https://doi.org/10.1002/ps.6048
Trenkamp S, Eckes P, Busch M, Fernie AR. 2009. 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* **5**:277–291. DOI: https://doi.org/10.1007/s11306-008-0149-8, PMID: 19718268
Van Etten M, Lee KM, Chang SM, Baucom RS. 2019. Parallel and Nonparallel Genomic Responses Contribute to Herbicide Resistance in Ipomoea Purpurea, a Common Agricultural Weed. *bioRxiv*. DOI: https://doi.org/10.1101/647164
Van Etten M, Soble A, Baucom RS. 2020. Variable Inbreeding Depression May Explain Associations between the Mating System and Herbicide Resistance. Cold Spring Harbor Laboratory. DOI: https://doi.org/10.1101/2020.09.25.313940
Varanasi VK, Brabham C, Norsworthy JK, Nie H, Young BG, Houston M, Barber T, Scott RC. 2017. A Statewide Survey of PPO-Inhibitor Resistance and the Prevalent Target-Site Mechanisms in Palmer amaranth (Amaranthus palmeri) Accessions from Arkansas. *Weed Science* **66**:149–158. DOI: https://doi.org/10.1017/wsc.2017.68
Vila-Aiub MM, Neve P, Powles SB. 2009. Fitness costs associated with evolved herbicide resistance alleles in plants. *The New Phytologist* **184**:751–767. DOI: https://doi.org/10.1111/j.1469-8137.2009.03055.x, PMID: 19825013
Vila-Aiub MM. 2019. Fitness of Herbicide-Resistant Weeds: Current Knowledge and Implications for Management. *Plants* **8**:469. DOI: https://doi.org/10.3390/plants8110469
Weng ML, Becker C, Hildebrandt J, Neumann M, Rutter MT, Shaw RG, Weigel D, Fenster CB. 2019. Fine-Grained Analysis of Spontaneous Mutation Spectrum and Frequency in *Arabidopsis thaliana*. *Genetics* **211**:703–714. DOI: https://doi.org/10.1534/genetics.118.301721, PMID: 30514707
Whaley CM, Wilson HP, Westwood JH. 2004. Characterization of a new ALS-inhibitor resistance mutation from the ALS gene of smooth pigweed (Amaranthus hybridus). *Weed Science Society of America* **54**:828–832. DOI: https://doi.org/10.1614/WS-05-040R.1
Whitcomb CE. 1999. An introduction to ALS-inhibiting herbicides. *Toxicology and Industrial Health* **15**:231–239. DOI: https://doi.org/10.1191/074823399678846592, PMID: 10188205
Yu Q, Powles SB. 2014. Resistance to AHAS inhibitor herbicides: current understanding. *Pest Management Science* **70**:1340–1350. DOI: https://doi.org/10.1002/ps.3710, PMID: 24338926
Yu G. 2020. Using ggtree to Visualize Data on Tree-Like Structures. *Current Protocols in Bioinformatics* **69**:e96. DOI: https://doi.org/10.1002/cpbi.96, PMID: 32162851
Zhao L, Deng L, Zhang Q, Jing X, Ma M, Yi B, Wen J, Ma C, Tu J, Fu T, Shen J. 2018. Autophagy contributes to sulfonylurea herbicide tolerance via GCN2-independent regulation of amino acid homeostasis. *Autophagy* **14**:702–714. DOI: https://doi.org/10.1080/15548627.2017.1407888, PMID: 29377765
Zhao LX, Jiang MJ, Hu JJ, Zou YL, Gao S, Fu Y, Ye F. 2020. Herbicidal activity and molecular docking study of novel PPO inhibitors. *Weed Science* **68**:565–574. DOI: https://doi.org/10.1017/wsc.2020.66
Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**:3326–3328. DOI: https://doi.org/10.1093/bioinformatics/bts606, PMID: 23060615
Zulet A, Gil-Monreal M, Villamor JG, Zabalza A, van der Hoorn RAL, Royuela M. 2013. Proteolytic pathways induced by herbicides that inhibit amino acid biosynthesis. *PLOS ONE* **8**:e73847. DOI: https://doi.org/10.1371/journal.pone.0073847, PMID: 24040092

doi:10.1111/evo.14404

Selective ancestral sorting and de novo evolution in the agricultural invasion of *Amaranthus tuberculatus*

Julia M. Kreiner,[1,2,3,4] Amalia Caballero,[5] Stephen I. Wright,[1,*] and John R. Stinchcombe[1,6,*,†]

[1]*Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON V6T 1Z4, Canada*

[2]*Current Address: Department of Botany, University of British Columbia, Vancouver, BC V6T 1Z4, Canada*

[3]*Current Address: Biodiversity Research Centre, University of British Columbia, Vancouver, BC V6T 1Z4, Canada*

[4]*E-mail: julia.kreiner@ubc.ca*

[5]*Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada*

[6]*Koffler Scientific Reserve, University of Toronto, King City, ON L7B 1K5, Canada*

Received August 3, 2021
Accepted October 24, 2021

The relative role of hybridization, de novo evolution, and standing variation in weed adaptation to agricultural environments is largely unknown. In *Amaranthus tuberculatus*, a widespread North American agricultural weed, adaptation is likely influenced by recent secondary contact and admixture of two previously isolated lineages. We characterized the extent of adaptation and phenotypic differentiation accompanying the spread of *A. tuberculatus* into agricultural environments and the contribution of ancestral divergence. We generated phenotypic and whole-genome sequence data from a manipulative common garden experiment, using paired samples from natural and agricultural populations. We found strong latitudinal, longitudinal, and sex differentiation in phenotypes, and subtle differences among agricultural and natural environments that were further resolved with ancestry inference. The transition into agricultural environments has favored southwestern var. *rudis* ancestry that leads to higher biomass and treatment-specific phenotypes: increased biomass and earlier flowering under reduced water availability, and reduced plasticity in fitness-related traits. We also detected de novo adaptation in individuals from agricultural habitats independent of ancestry effects, including marginally higher biomass, later flowering, and treatment-dependent divergence in time to germination. Therefore, the invasion of *A. tuberculatus* into agricultural environments has drawn on adaptive variation across multiple timescales—through both preadaptation via the preferential sorting of var. *rudis* ancestry and de novo local adaptation.

KEY WORDS: De novo adaptation, gene flow, phenotypic plasticity, preadaptation, weed evolution.

Although selection from herbicides is one of the most dramatic and novel selection pressures that new agricultural weed populations experience, a much broader suite of ecological shifts and adaptive changes is likely to accompany the transition into agronomic environments (Murphy and Lemerle 2006) and resulting range expansion (Clements and Ditommaso 2011). Weeds that have successfully invaded contemporary landscapes, including crop fields and range lands, are subject to predictable and repeated disturbances, regimented irrigation, extreme interspecific competition, and intensified chemical inputs—all of which should lead to novel selection pressures to accelerate life history and assure reproduction in variable environments (Baker 1974; De Wet and Harlan 1975; Vigueira et al. 2013). Baker (1974) hypothesized that in addition to specialized traits, an "ideal weed" might possess a phenotypically plastic generalist genotype to better respond to agricultural disturbance regimes. Because of the impact of weed populations on crop productivity and native

*Centre for the Analysis of Genome Evolution and Function,University of Toronto,Toronto,ON,M5S 1A8,Canada

†These authors contributed equally to this work.

diversity, agricultural weeds present a particularly pressing case study of convergent adaptation across species, yet remain relatively neglected in the field of evolutionary genetics (Stewart et al. 2009; Ravet et al. 2018; Martin et al. 2019). Indeed, despite long standing hypotheses of the direction of weed evolution (De Wet and Harlan 1975) and caricatures of ideal weeds (Baker 1974), the phenotypic changes that result from the transition from natural to agricultural environments, as well as the origins and relevant timescale of genetic variation that underlies these changes, remain unresolved in most systems (but see Barrett 1983; Boudry et al. 1993; Arnaud et al. 2010; Muller et al. 2011; Kuester et al. 2014; Charbonneau et al. 2018; Ye et al. 2019).

Identifying the genetic source of weed adaptation—whether arising de novo, drawing from standing variation, or through gene flow—has important implications for understanding the tempo of evolution and inevitability of weed persistence in agricultural settings. The role of gene flow in weed adaptation to agriculture has been well-recognized, especially via hybridization of wild and domesticated relatives (De Wet and Harlan 1975). Hybrid origins of invasive weed populations have been well-documented in the genus *Helianthus* (Kane and Rieseberg 2008; Muller et al. 2011; Lai et al. 2012), with multiple wild to weedy transitions occurring via crop hybridization. In wild and cultivated beets (*Beta vulgaris*), hybridization has led to invasive weed populations with a mix of agriculturally fit traits of both types, including self-fertilization (from the domesticated type), early bolting, and annual flowering (wild type traits) (Arnaud et al. 2010). Thus, although hybridization of weeds with domesticates may act as a direct line to adaptive genetic variation, gene flow between locally adapted types within a species ("ecotypes," sensu Turesson 1922), common in many weeds (Brown and Marshall 1981; Barrett 1982), may further facilitate a rapid response to selection (Fisher 1930; Baker 1974). The role of within-species standing variation (related to "preadaptation," sensu Liebman et al. 2001, referring to prior adaptation leading to high fitness in a novel environment) versus de novo evolution is largely untested in agricultural weeds, although it has been investigated more extensively in invasive alien plant species (e.g., Guo et al., 2014; Schlaepfer et al., 2010).

Amaranthus tuberculatus is a diploid annual native to North America (Costea et al. 2005), the genetics of which is highly diverse and geographically structured (Waselkov and Olsen 2014; Kreiner et al. 2019). It is extremely successful in agricultural systems, hypothesized to result in part from a combination of its obligately outcrossing dioecious wind-pollinated mating system (Costea et al. 2005) and extremely high seed production (with females producing on average between 35,000 and 1,200,000 notably small [1 mm] seeds [Stevens 1932; Sellers et al. 2003; Hartzler et al. 2004]). Recent inference in the species highlights a massive recent expansion in effective population size over the last century—a key consequence of which is highly parallel target-site resistance evolution (Kreiner et al. 2021). In *A. tuberculatus*, two major lineages and ecotypes exist, the classification of which has been debated and revised from two species (Riddell 1835; Sauer 1955) to one (Uline and Bray. 1895; Pratt and Clark 2001), to most recently, two distinct varieties on the basis of continuous, clinal morphological variation across their sympatric ranges (Costea and Tardif 2003; Costea et al. 2005). We will refer to these lineages as varieties throughout.

The two *A. tuberculatus* varieties differ in their historical ranges as inferred from herbarium specimens, with var. *tuberculatus* being found along northeastern Missouri and Mississippi water basins, but var. *rudis* (initially circumscribed as *A. tamariscinus*) historically restricted to ruderal habitats in four southwestern states in the United States (Sauer 1957). The secondary contact of these varieties over the last two centuries was thought to be driven predominantly by the expansion of var. *rudis* northeastwards. Sauer (1957) hypothesized that the admixture resulting from this secondary contact led to the agriculturally competitive form. However, he also posited that the hygrophytic nature of species in the genus, and their conditioning to frequently disturbed riparian habitats, preadapted them to the human-mediated disturbances widespread in agricultural landscapes. Recent genetic and genomic evidence of a longitudinal cline in ancestry between their ancestral ranges supports the secondary contact of *A. tuberculatus* varieties, but the tendency to see var. *rudis* ancestry in agricultural environments (Waselkov and Olsen 2014; Kreiner et al. 2019) suggests that var. *rudis*, in particular, may be preadapted.

Although differences in ecological pressures across natural and agricultural habitats may shape patterns of phenotypic and genomic diversity, this fine-scale evolution is likely to be mediated by geographic gradients in abiotic factors that determine seasonality across broader scales. Adaptive geographic clines in plant traits, latitudinal clines in particular, are ubiquitous and have been widely described across systems for reproductive, defense, and growth-related phenotypes (Neuffer 1990; Stinchcombe et al. 2004; Samis et al. 2012; Peterson et al. 2016; Cornille et al. 2018; Bilinski et al. 2018; Frachon et al. 2018; Exposito-Alonso 2020). In short day plants (i.e., those where flowering is induced by the shortening of days at the end of the growing season), individuals at higher latitudes should be selected to flower earlier to set seed before frost-induced mortality (Holm 2010). Longitudinal clines are less common but have been described for flowering time in *Arabidopsis* where it is thought to be associated with geographic variation in winter temperature and precipitation (Samis et al. 2008, 2012). Given the latitudinal and longitudinal variation in *A. tuberculatus* ancestry, adaptation

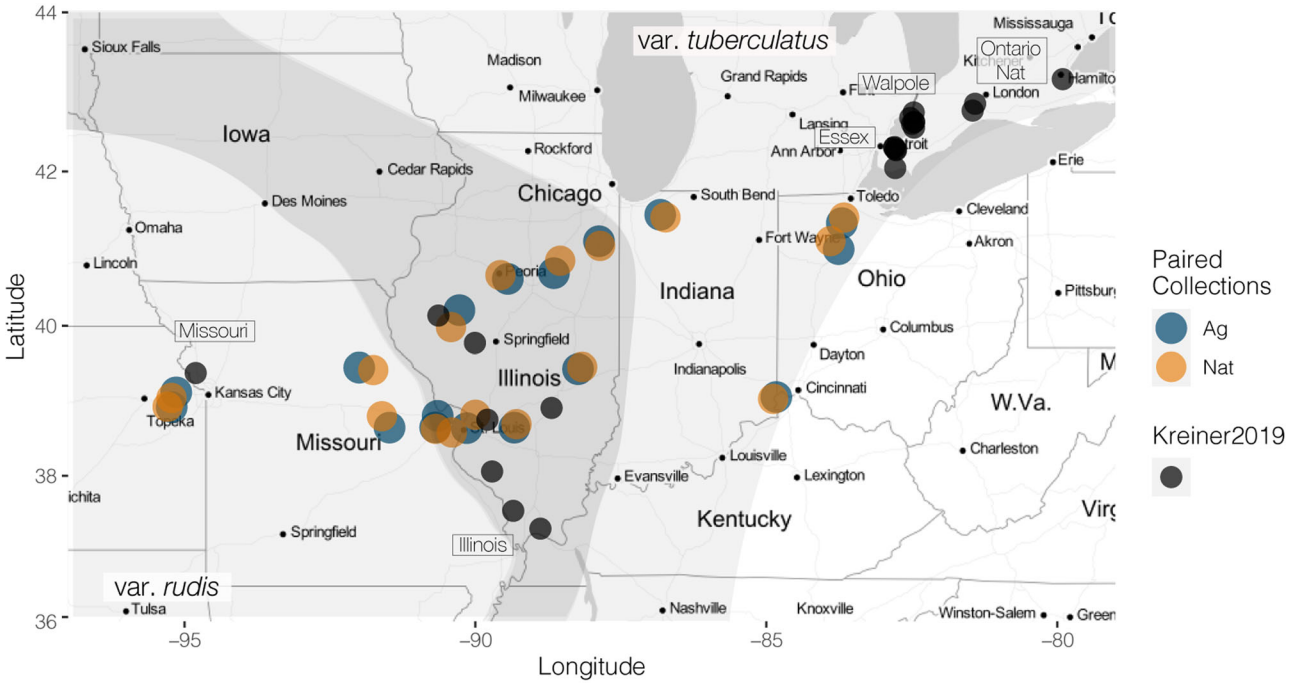

Figure 1. **Pairwise collections of natural and agricultural populations spanning the historical sympatric (dark shaded area) and allopatric ranges of *A. tuberculatus* var. *tuberculatus* (northeast) and *A. tuberculatus* var. *rudis* (southwest; range limits adapted from Sauer [1957]). For context, we also depict populations along with their regional label from Kreiner et al. (2019) in black.**

to these climate gradients may be in part confounded with historical patterns of ancestral divergence.

Here, we test key hypotheses about the role of admixture, de novo, and ancestral variation in facilitating the recent invasion of *A. tuberculatus* into agricultural environments. A recent study performed a replicated common garden experiment using a broad collection of *A. tuberculatus* to test hypotheses about agricultural adaptation, but was largely unable to uncouple geographic and fine-scale environmental drivers of phenotypic differentiation (Waselkov et al. 2020). To ensure sufficient power to disentangle broad geographic and environmental drivers of adaptation, we used a paired collection design (Lotterhos and Whitlock 2015), sampling *A. tuberculatus* in pairs of natural and agricultural sites that were <25 km apart, in a replicated fashion across 3 degrees of latitude and 12 degrees of longitude. We then tested for local adaptation to agricultural environments in a common garden experiment with treatments simulating components of natural and agricultural environments. We performed a water-supplemented treatment to simulate a key component of the riparian habitats in which natural populations were collected, and a soybean (*Glycine max*) competition treatment that was the predominant focal crop where agricultural *A. tuberculatus* was collected. Lastly, we implemented a control treatment that lacked both competition and water supplementation. Across collections from 17 sets of paired natural and agricultural populations (34 populations in total), we grew 10 replicates of full siblings from 200 maternal lines across each of the three treatments, totaling to 6000 individuals. Key to testing hypotheses about the timescale of adaptation, we also collected whole genome sequence data from 187 maternal lines to explicitly examine the extent to which ancestry drives phenotypic differentiation across natural and agricultural environments and geographic clines. By combining a paired sampling approach, a highly replicated phenotypic catalogue, and genomic data, our results provide robust insight into the impact of human-mediated disturbances on trait differentiation and the timescale underlying adaptation to contemporary agricultural environments.

Methods

COLLECTIONS AND PARENTAL CROSSES

We made collections of 17 paired populations (a natural and agricultural population collected <25 km apart, 34 populations in total) in October 2018, from Ohio to Kansas, aiming for 20 maternal lines per population and ranging from eight to 30 maternal lines sampled (Fig. 1). Seed was partitioned into mesh jewelry bags and buried in moist sand for 6 weeks before being grown out, as per stratification recommendations for the species (Leon et al. 2007). Four replicates of 700 maternal lines across these populations were sown and grown in growth chambers, under short day conditions to shorten generation time, and germinated under a 12-degree temperature amplitude to maximize germination (Leon et al. 2004) (16 h at 32 degrees, 8 h at

20 degrees). Upon formation of reproductive organs, females and males were immediately bagged to prevent cross-pollination (inspired by McGoey et al. 2017) until enough individuals had flowered that controlled, within-population crosses could be conducted. We conducted 345 within-population crosses, where we randomly assigned males to be transplanted into a pot of a female from a different maternal line within the same population, such that we maximized the number of crosses within populations while only performing one cross per maternal line. Upon transplanting the male into the female pot, we bagged the entire aboveground portion of the pot and agitated the bags to facilitate male pollen dehiscence. Seeds successfully set in 326/345 crosses, and were harvested for cold treatment prior to the common garden experiment.

ROOFTOP COMMON GARDEN EXPERIMENT

We subsampled 200 of the 326 F1 lines from within-population crosses, with the aim of matching sample size across natural and agricultural environments within each population pair. We cold-treated these lines in a 4°C growth chamber in the dark for 8 weeks, in 8-cm wide petri dishes with 7.5 ml of deionized water. We sowed seeds in 1-L treepots (Stuewe and Sons., Inc) in the greenhouse (3 July 2019), and initially grew them under fluctuating temperatures of 14°C at night (8 h) and 32°C in the day (14 h) to maximize germination. Soybean seed (*Glycine max* var. *dekalb*—DKB, 12–57) was sown in competition pots the next day, with A*maranthus* and soy equally spaced within the pot. Treepots assigned to the water treatment had their bottoms duct taped off to increase water retention. Plants were watered and checked for germination daily for 10 days, and then moved outside on to the roof (12 July 2019) into a fully randomized, complete block design, with every block (10) serving as a replicate of each maternal line (200) in every treatment (3), totaling to 6000 individuals. We thought it important to rear plants outdoors to expose them to natural temperature, rainfall, light, and various other environmental signals that may affect phenotypes. We should clarify that in contrast to classic reciprocal common garden experiment, reference to "Environment" as a predictor throughout this study refers to the source of collected genotypes (Natural habitats or Agricultural fields) and is distinct from the reared environment that we call "Treatment" (i.e., whether genotypes were grown in the Water, Control, or Soy treatment).

PHENOTYPING AND DATA COLLECTION

Two weeks after germination and 1 week after plants were moved onto the roof, we commenced phenotypic measurements starting with cotyledon width (mm), hypocotyl length (mm), and leaf number. With 6000 plants, this took about 10 days to complete, and so we also recorded date of measurement as a covariate to be used in related analyses. Once the first individual was found in flower, we checked all plants for the start of flowering Monday, Wednesday, and Friday for 4 weeks. Upon flowering, we also measured stem width, plant height, number of nodes, whether an individual was recorded late (extended inflorescence), or whether the plant had been damaged (these individuals were subsequently excluded from the statistical analyses). Due to a long tail of flowering, after 4 weeks we halved census efforts, alternately checking half of the blocks each Monday, Wednesday, and Friday. Above ground biomass for all undamaged plants was harvested into paper bags, starting 8 weeks after the start of flower and lasting until 11 weeks after flowering until all plants had been harvested. Upon harvest, we recorded date, sex, flower color, and stem color. Plants were then dried in a 50°C oven for 3 days and weighed for above ground biomass. In total, we measured 11 phenotypes of interest: days until germination, cotyledon width, hypocotyl height, early leaf number, time to flowering, height at flowering, node number at flowering, stem width at flowering, flower color (visual rating on a scale of 1 [light green] to 4 [dark purple]), stem color (visual rating on a scale of 1 [light green] to 4 [dark purple]), and dry biomass. We also tracked sex, greenhouse number, greenhouse block, roof block, days to first measurement, and days to harvest. For visualization, occasionally phenotypic rates are shown, which are calculated by dividing the focal trait by the number of days between measurement and germination.

DNA COLLECTIONS AND SEQUENCING

We sampled two to three of the youngest leaves on each individual in two blocks of our common garden experiment just before harvest. Because each block contains replicates of the same family lines, we chose to sample these two blocks to have backup tissue for each family in each treatment (backups that were later destroyed by a −80°C freezer failure during COVID-19). Leaves were immediately put in tubes and submerged in liquid nitrogen before being stored at −80°C until extraction. We extracted DNA from the 200 unique maternal lines that were grown in the common garden experiment. Total DNA was extracted using Qiagen DNeasy Plant Mini kit according to manufacturer's instructions. We sent DNA samples to Genome Quebec Innovation Centre (McGill University), Montréal, QC, Canada for library preparation and sequencing; 187 ended up being sequenced due to extraction and library quality. Libraries were prepared using the NEB Ultra II Shotgun gDNAlibrary preparation method and sequenced on four lanes of Illumina NovaSeq S4 PE150 (2 × 150) sequencing platform using 96 barcodes. A total of ∼25 billion reads (25,818,840,892) were generated, with an average of ∼137 million (137,334,200) per individual.

MAPPING AND SNP CALLING

We aligned reads to the female *A. tuberculatus* reference genome (Kreiner et al. 2019), using BWA-mem version 0.7.17-r1188

(Li 2013). After mapping, individuals had an average diploid coverage of 28×. Duplicate reads were removed with picard MarkDuplicates (Broad-Institute 2016). We used Freebayes version 1.1.0-46 (Garrison 2012) to call SNPs using default settings except for –max-complex-gap 1, –haplotype-length 1, and –report-monomorphic. We then filtered SNPs such that sites were removed based on excess missing data (>20%), allelic bias (AB <0.25 and >0.75), overall variant call quality (QUAL <30, removing sites with greater than a 1/1000 SNP calling error rate), after dustmasking for low complexity and removing multiallelic SNPs and indels. Because high coverage data tend to overestimate mapping quality such that it no longer scales linearly with depth, we followed recommendations in Fang (2014), further removing particularly high depth sites (>mean depth + 3(sqrt(mean depth))) with relatively low QUAL (<2*depth). Five genotypes were removed from downstream analyses due to >5% sequencing error rate based on a KMER-based analysis (Ranallo-Benavidez et al. 2020), resulting in a total of 20,555,154 high-quality SNPs across 182 individuals.

POPULATION STRUCTURE

We merged filtered, high-quality SNPs (using the program bcftools merge) from the 182 high-quality resequenced genomes described above, with the high-quality SNP set from Kreiner et al. (2019). Important for merging these datasets, the same SNP filtering process was applied to both datasets, except for the high-coverage step, as the Kreiner et al. (2019) collections were of more moderate coverage (∼10×). Briefly, these previous collections were made from eight agricultural fields in Illinois and Kansas with reports of uncontrolled *A. tuberculatus* popuations, and fields within Walpole Island and Essex County, Ontario, Canada, where reports of agriculturally associated populations of *A. tuberculatus* have only recently occurred in the last decade. Additionally, this dataset included 10 individuals collected from nearby Ontario natural populations, occurring alongside the Thames River outside of London, and the Grand River outside of Hamilton. From this merged set of 2,591,759 SNPs, we investigated population structure with a principal component analysis (PCA) in plink (option –pca) (Purcell et al. 2007). We investigated predictors of genome-wide relatedness in a multiple regression framework using principal component (PC) 1 and PC2 as dependent variables and longitude, latitude, sex, environment, and population pair as independent variables. To test if predictors were different among PCs, we used a grouped regression approach that included values of both PCs at once, testing whether the explanatory power of these various predictors differed among PCs (as indicated by an interaction with principal component number, i.e., PC1 or PC2) (model syntax: PC value ∼ PC # × Environment + PC # × Longitude + PC # × Sex + PC # × Pair). Lastly, we used the program Faststructure (Raj et al. 2014) to estimate the proportion of individual and population admixture levels, at $K = 2$ (testing a priori hypotheses about the distribution of *A. tuberculatus* varieties) and for comparison, at $K = 3$.

MIXED MODELS AND TESTS FOR PREADAPTATION

Modeling individual-level phenotypic variation

We used R to fit linear mixed models (implemented with the package lme4) of geographic, environmental, and sex-based predictors to each of the eight quantitative phenotypes measured in the common garden experiment, all of which were evaluated with a type III sums of squares. For the two categorical phenotypes (flower color, stem color), we used the R package glmmadmb to implement a multinomial mixed model (link = "logit"). For analysis of phenotypic variation at the individual level from the common garden experiment, we accounted for the relevant block effect (typically roof block, except for time to germination, for which we used greenhouse: greenhouse block), and nested hierarchical structure of maternal lines (family) within populations as random effects. Because we were testing independent hypotheses about factors that drive variation across our different phenotypes (and thus testing each hypothesis once), no multiple test correction was performed. For all 11 traits, individual-level phenotypic variation was measured with the following model structure (note that we initially included an environment by treatment interaction but removed it due to low explanatory power and lack of significance for all phenotypes, except for germination):

$$\begin{aligned}\text{Focal trait} \sim\ & \text{Environment} + \text{Treatment} + \text{Sex} \\ & + \text{Lat} + \text{Long} + \text{Germination JD} + \text{Measurement JD} \\ & + \text{Population mean ancestry} + (1|\text{Block}) + (1|\text{Pop/Family})).\end{aligned} \quad (1)$$

Our fixed effect predictors had the following characteristics: environment had two levels (natural or agricultural), treatment had three levels (control, soy, or water), sex had two levels (M/F), and latitude and longitude were both treated as continuous variables, referring to the geographic coordinate of the originating population. Except for when we were modeling germination itself, Julian day of germination and Julian day of measurement were included as covariates to account for variation in how long a plant had been growing prior to measurement. We used the population-mean varietal ancestry (the average of the Faststructure inferred proportion of an individual's genome assigned to cluster 1 at $K = 2$ across all sequenced individuals within a population) as an estimate of genetic structure in these models. We used population-level estimates rather than family level because of low sequencing replication (one individual per full sibling family in an obligately outbreeding species), and because population-level estimates should reflect broad-scale geographic patterns in ancestry, similar to using latitude and

longitude as proxies for geographic and climatic variation. Using Akaike information criterion (AIC), we compared the full model as shown above to a reduced model that excluded population mean ancestry to evaluate its importance in explaining phenotypic differentiation.

Testing the role of var. rudis *ancestry*

We were interested in explicitly testing the role of var. *rudis* ancestry on adaptive phenotypic variation, given a priori alternative hypotheses of hybridization versus var. *rudis* ancestry facilitating agricultural adaptation (Sauer 1957; Waselkov 2014). To incorporate both ancestry and phenotypic traits, we analyzed the sex-specific phenotypic means within each treatment, within each maternal line. Thus, while phenotypic family means were distinct across each sex and treatment level with a maternal line, we assigned all treatment and sex replicates of a maternal line the Faststructure ancestry estimate we attained from their single sequenced full sibling. We then examined two key fitness-related traits, biomass and flowering time, separately for males and females given strong sexual dimorphism in the species.

Beyond the direct linear effect that var. *rudis* ancestry might have on phenotypic variation, we were particularly interested in testing whether there was a positive quadratic effect on fitness-related variation (indicating a role for heterosis), and whether there was an interaction between the proportion of var. *rudis* ancestry and experimental treatment on fitness-related variation (indicating a role for preadaptation, in that ancestry confers reared environment-specific benefits). Lastly, we wanted to test the extent to which phenotypes differed among natural and agricultural environments, regardless of ancestry (indicating de novo agricultural adaptation). To test these three alternative but not mutually exclusive hypotheses, we used a model similar to one testing linear selection gradients separately for males and females, regressing these ancestry and environment terms alongside standardized phenotypic predictors on two fitness-related traits (flowering time and biomass) to account for correlated trait evolution:

$$\begin{aligned}\text{Fitness} - \textit{related}\ \text{trait} \sim\ & \text{stand (Germination time)} \\ & + \text{stand}\left(\text{Hypocotyl length}\right) + \text{stand}\left(\text{Cotyledon width}\right) \\ & + \text{stand (Leaf number)} + \text{stand (Stem width @ FT)} \\ & + \text{stand}\left(\text{Plant height @ FT}\right) + \text{stand (Node number @ FT)} \\ & + \text{stand}\left(\text{Flowering time or biomass}\right) + \text{stand (Stem color)} \\ & + \text{stand (Flower color)} + \text{Long} + \text{Lat} + \text{Treatment} + \mathbf{Env} \\ & + \mathbf{Ancestry} + \mathbf{Ancestry^2} + \mathbf{Ancestry} : \mathbf{Treatment}. \qquad (2)\end{aligned}$$

Results

DRIVERS OF THE DISTRIBUTION OF ANCESTRAL VARIATION

To understand patterns of genetic relatedness that underlie phenotypic variation within and between populations, and the potential role of ancestry in facilitating agricultural adaptation, we first characterized patterns of population structure and ancestry across our accessions in the context of previously studied populations (Fig. 1).

We find that individuals from our paired-environment collections show a longitudinal cline in ancestry, as expected (Sauer 1957; Waselkov and Olsen 2014; Kreiner et al. 2019) (Fig. 2A, C). Previous work showed that Ontario natural populations in the eastern part of the range are homogenous for var. *tuberculatus* ancestry and that, along with our most westerly collections in Missouri, nearby Essex county agricultural populations are homogenous for var. *rudis* ancestry, likely reflecting a long-distance introduction event from the Midwest (Kreiner et al. 2019). The nearly 200 genotypes we have added to this genome-wide inference of population structure support the circumscription of the historical ranges of these two ancestral varieties, in that northeastern populations (e.g., Maume, Mccombe, Weston) showed a higher proportion of *Amaranthus* var. *tuberculatus* ancestry and southwestern populations showed predominantly var. *rudis* ancestry (Fig. 2C). A joint PCA of genome-wide genotypes from common garden accessions and samples previously characterized in (Kreiner et al. 2019) illustrates that our newly collected accessions showed somewhat less extreme population structure along both the first and second PCs. This is consistent with the more continuous but geographically intermediate sampling we performed for genotypes phenotyped and sequenced in the common garden (Figs. 1 and 2A). Individuals from our common garden typically fell in a very similar position for PC2 and showed much more variation along PC1, which explained 20% of the total variation in genotype composition across the 349 joint accessions. PC1 has been previously shown to strongly reflect *A. tuberculatus* varietal ancestry (Kreiner et al. 2019).

When we performed a PCA exclusively on genotypes from our common garden experiment, PC1 similarly explained 18% of the variation in genotype composition, whereas PC2 explained substantially more than the joint PCA, 15% (Fig. 2B). From a multivariate regression of this PCA of just common garden accessions, we find that PC1 significantly relates to longitude ($F_{1,180} = 27.05$, $P < 0.001$), population pair ($F_{1,180} = 4.58$, $P = 0.03$), and environment (agricultural vs. nonagricultural, $F_{1,180} = 5.51$, $P = 0.02$), but neither latitude nor sex. To test if the predictive effects of environment differ across PCs in a joint model, we performed a follow-up grouped regression, jointly examining if the predictive

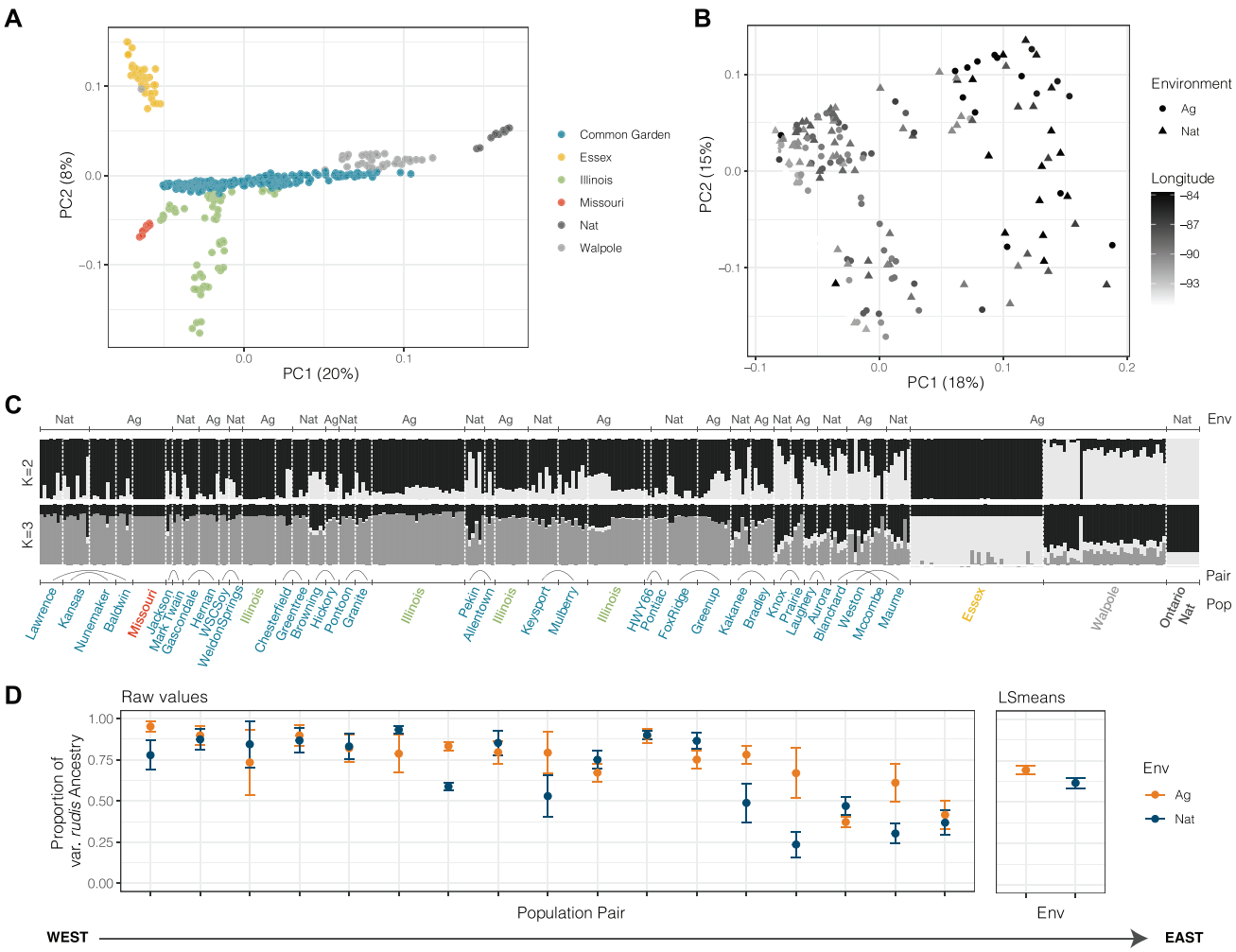

Figure 2. **Patterns of population structure from paired natural-agricultural collections, grown in the common garden experiment, in the context of samples from Kreiner et al. (2019). (A) Principal component analysis of samples from Kreiner et al. (2019), where populations have been identified as homogenous for both *A. tuberculatus* var. *tuberculatus* ancestry (e.g., Ontario Nat) and *A. tuberculatus* var. *rudis* ancestry (e.g., Missouri) along with 187 genotypes grown in the common garden experiment originating from pairwise Nat-Ag population sampling. (B) Principal component analysis of just common garden genotypes. (C) Population structure (at $K = 2$, reflecting ancestry of var. *rudis* in black and var. *tuberculatus* in light gray) across both previously analyzed samples and common garden accessions. Allowing for another ancestral population ($K = 3$) reflects largely the same major axes of variation in $K = 2$. Plot is sorted by longitude (from west to east), population pairs within the common garden experiment are indicated by arched lines, and labels of populations from Kreiner et al. (2019) are colored according to the legend in panel A. (D) Higher proportion of average var. *rudis* ancestry in agricultural versus natural environments within population pairs, sorted by longitude (left), and the average effect across environments as illustrated by the least-squares means from a multiple regression that also included longitude, latitude, and pair (right). Error bars represent standard error.**

effects of environment (agricultural vs. natural), pair, and longitude differ among PCs (i.e., testing for a significant PC number × predictor interaction). This grouped model fails to detect a significant environment by PC interaction, implying that the predictive effects of environment are consistent across multiple dimensions of genotype differentiation, but picks up a significant pair × PC ($F_{1,366} = 10.3686$, $P = 0.001396$) and longitude × PC interaction ($F_{1,366} = 84.95$ $P = <2.2 \times 10^{-16}$) with both pair and longitude better predicting the first PC.

The influence of environment (whether genotypes were collected in natural or agricultural environments) on ancestry identified in the common garden-specific PCA is apparent in Figure 2C, where agricultural populations show an excess of var. *rudis* ancestry given their longitude, and more apparently so in Figure 2D, within their population pair—a more direct comparison of environmentally driven sorting of ancestry. Indeed, a multivariate regression of the Faststructure-inferred proportion of var. *rudis* ancestry (the proportion of grouping 1 at $K = 2$) finds

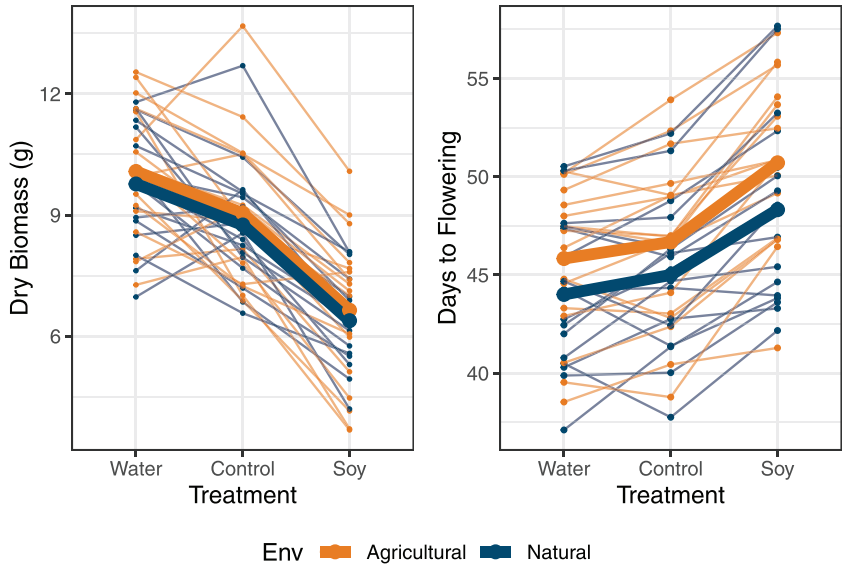

Figure 3. Population-level reaction norms of biomass and flowering time across the water, control, and soy treatments in the common garden experiment. Thin and thick lines represent population and environment mean reaction norms, respectively, and are additionally colored by whether collections were found in natural or agricultural environments.

longitude ($F_{1,167} = 8.29$, $P = 0.005$), pair ($F_{1,167} = 3.25$, $P = 6.70 \times 10^{-5}$), and environment ($F_{1,167} = 6.66$, $P = 0.011$) to be significant predictors of ancestry, with more var. *rudis* ancestry in agricultural environments. On average, this pattern resulted in a 7.8% excess of var. *rudis* ancestry in agricultural environments across all population pairs, after controlling for other covariates. Of our 17 population pairs, the nine pairs that show greater var. *rudis* ancestry in agricultural environments have a median 25% excess of var. *rudis* ancestry compared to natural environments (and up to 43% greater in the most extreme pairing; Fig. 2D). In contrast, the remaining eight pairs that have greater var. *rudis* ancestry in natural environments differ by only 8% on average (and at 14% at its maximum). The significant enrichment of var. *rudis* ancestry in agricultural environments given a population pair's longitude supports the hypothesis that the expansion of the var. *rudis* contributed to the *A. tuberculatus* agricultural invasion (Waselkov and Olsen 2014; Waselkov et al. 2020). Furthermore, that population pair significantly predicts ancestry across a disparate sampling suggests that selection is maintaining this pattern of environment-dependent ancestry despite nearby natural and agricultural populations being highly connected through gene flow.

GENERAL OBSERVATIONS AND PLASTICITY IN A MANIPULATIVE COMMON GARDEN

Across the 4493 individuals fully phenotyped in the common garden experiment, we found almost a perfect 1:1 sex ratio (2252 males vs. 2241 females). On average, we completely phenotyped 22.5 families per population with 11.2 females per family (SD = 3.17) and 11.3 males per family (SD = 3.33), with an average of 7.5 family replicates phenotyped across each of three treatments.

Our treatments worked as expected, with population-mean flowering time and dry biomass reflecting that plants generally grew larger and flowered fastest in the water treatment, and were smallest and later flowering in the soy competition treatment (Fig. 3). However, rather than recapitulating the often-flooded environment of natural populations, from eye observations and the necessity of near daily watering, the water treatment only tended to reduce drought stress relative to the control treatment. As an example of the magnitude of the effect of our three treatments, we characterized how flowering time differed depending on whether a genotype was reared in the water, control, or soy treatment. A least-squared mean estimate from the general regression model, using flowering time as a response variable, estimates that the water treatment led to 1 day earlier flowering than the control and 4 days earlier than the soy treatment. We further modeled phenotypic plasticity as a random effect, testing for a family by treatment interaction (lmer notation = 1 | treatment:family). Following the general model additionally including this plasticity random effect term, we find that modeling phenotypic plasticity results in a significantly better fit to our data despite the additional degrees of freedom ($\chi^2_{df=1} = 369.27$, $P < 0.001$) and can explain an additional 5.28% of the variation in flowering time (on top of the 51% of the base model; conditional r^2)—implying 10% (0.053/0.51) of the explainable variation in flowering time is plastic. Phenotypic plasticity was of even greater importance for determining dry biomass, explaining an additional 11% of variation in dry biomass on top of the 64% that can be explained in

our base model ($\chi^2_{df=1} = 320.3824$, $P = <2 \times 10^{-16}$), implying that ~17% (0.11/0.64) of the explainable variation in above ground biomass is plastic. To test whether populations from natural and agricultural environments differed in the extent of phenotypic plasticity, we compared our plasticity model to one that allowed plasticity to differ across environments (lmer notation = Environment | Treatment:Family). For both biomass and flowering time, allowing plasticity to vary among environments did not increase the variance explained in the model, with environmental differences in plasticity not significantly explaining dry biomass ($\chi^2_{df=2} = 0.329$, $P = 0.849$) and very marginal effects on flowering time ($\chi^2_{df=2} = 4.613$, $P = 0.0996$) (Fig. 3).

DRIVERS OF PHENOTYPIC VARIATION AND THE ROLE OF ANCESTRY

Geographic, environmental, and ancestral trait divergence

We found evidence for phenotypic differentiation across our broad sampling of *A. tuberculatus* individuals grown in the same common garden, by latitude, longitude, sex, and between agricultural and natural environments. We evaluated phenotypic variation across all individuals in the common garden experiment in the typical manner of controlling for nested family structure, but also considering the effect of accounting for ancestry. For all models, adding population-mean ancestry as a covariate led to a substantially smaller AIC (Table 1).

Despite sampling a far greater range of longitude than latitude, after accounting for ancestry, latitude more consistently predicted variation in our measured traits (6/11 traits significantly predicted by latitude vs. 3/11 for longitude: Table 1). For longitude, accounting for ancestry tended to decrease its explanatory power, considerably decreasing the longitude χ^2 and significance for days to flowering, near completely so for stem color (Table 1; Fig. 4). Although latitude also covaried with ancestry, accounting for ancestry tended to increase the explanatory power of latitude (e.g., for stem width at flowering and dry biomass) (Table 1; Fig. 4). The observation that ancestry absorbs more explanatory power of longitude compared to latitude is consistent with stronger longitudinal than latitudinal isolation between *A. tuberculatus* lineages, in line with the patterns of population structure we describe above.

Amaranthus tuberculatus ancestry significantly predicted days to germination ($\chi^2 = 4.121$, $P = 0.043$), hypocotyl length ($\chi^2 = 9.687$, $P = 0.0019$), stem width at flowering ($\chi^2 = 13.824$, $P = 0.0002$), stem color ($\chi^2 = 29.572$, $P = 5.39 \times 10^{-8}$), flower color ($\chi^2 = 6.008$, $P = 0.0142$), and marginally dry biomass ($\chi^2 = 3.490$, $P = 0.0618$) in our individual-level regressions (Table 1)—highlighting the role of the historical isolation between these two lineages in shaping current day patterns of phenotypic variation from early to late-life history. We found little signal of the classic reciprocal common garden test for agricultural adaptation across our measured traits ("a home advantage"); however, days to germination showed a significant treatment by environment interaction ($\chi^2 = 9.376$, $P = 0.009$), with agricultural and natural types having similar time to germination in the control and soy treatment, but notably earlier germination of riparian natural types in the water treatment (Fig. 4). With genetic ancestry showing significant differences between the two habitats of origin, we examined the extent that accounting for ancestry resolved natural-agricultural phenotypic differentiation. In comparison to the full model results, where environment was a marginally significant predictor only for days to flowering ($\chi^2 = 3.038$, $P = 0.0814$), before accounting for ancestry, agricultural types tended to have marginally wider cotyledons ($\chi^2 = 2.866$, $P = 0.0905$), fewer leaves early on ($\chi^2 = 2.857$, $P = 0.090$), and showed significantly longer time to flowering ($\chi^2 = 4.376$, $P = 0.0365$) (Table 1; Fig. 4). The signficant environment by treatment interaction for time to germination was consistent across models (Full model: $\chi^2 = 9.3759$, $P = 0.0092$; Ancestry-reduced model: $\chi^2 = 9.498$, $P = 0.0087$).

Ancestry effects on fitness-related traits: Tests for preadaptation, de novo adaptation, and hybrid vigor

We hypothesized that if var. *rudis* ancestry is preadapted to agricultural habitats, the effect of the proportion of var. *rudis* ancestry on key life history characteristics such as biomass and flowering time would vary depending on experimental treatment. In contrast, if hybridization between varietal lineages has facilitated much of the *A. tuberculatus*'s contemporary invasion through hybrid vigor, we predicted that var. *rudis* ancestry would have a nonlinear relationship with fitness-related traits. Finally, if populations were adapting to agricultural regimes de novo, we predicted that fitness-related traits should vary among natural and agricultural environments, regardless of ancestry.

An analysis at the family-mean level of lifetime aboveground biomass in males found no quadratic effect of var. *rudis* ancestry; however, we found a significant linear effect of the proportion of var. *rudis* ancestry, where pure var. *rudis* types were predicted to accumulate biomass at a rate of 0.046 g/day more than pure var. *tuberculatus* types ($F_{1,498} = 3.9112$, $P = 0.0485153$). Additionally, the interaction effect of var. *rudis* ancestry by treatment and marginally, source environment (natural or agricultural), significantly affected male biomass (Ancestry × Treatment: $F_{2,498} = 3.335$, $P = 0.0364$; Environment: $F_{2,497} = 3.234$, $P = 0.0728$). Biomass-based fitness estimates tended to be lower in males from natural environments, compared to agricultural environments, regardless of treatment (Fig. 5). The significant interaction between var. *rudis* ancestry and treatment revealed that the proportion of var. *rudis* ancestry had little effect on biomass in the water and soy treatment, but that it substantially

Table 1. Summary of the results of linear or generalized mixed models for 11 focal traits measured across three life history stages, comparing the full model to a reduced model that drops population-mean ancestry as a covariate.

Early-life history traits	Significant predictors	ΔAIC (full-red.)	Predictors that change *P*-value class with + Ancestry ($\Delta\chi^2$)
Days to germination	Treatment (.), Env × Treatment (**), Sex (**), Anc (*)	−407.40	NA
Hypocotyl length	Treatment (***), Lat (***), Anc (**)	−515.50	NA
Cotyledon width	*NA*	−299.64	Environment (2.87→1.57)
Early leaf number	Treatment (***), Long (.), Lat (.), Sex (**)	−654.74	Environment (2.86 →1.41), Sex (6.47→7.29)
Mid-life history traits			
Height at flowering	Treatment (***), Lat (***), Sex (***)	−642.48	NA
Node number at flowering	Treatment (***), Long (*), Sex (***)	−484.63	NA
Stem width at flowering	Treatment (***), Lat (***), Sex (***), Anc (***)	−330.20	Latitude (5.39→13.91)
Days to flowering	Env (.), Treatment (***), Sex (*), Lat (**), Long (*)	−574.04	Longitude (8.07→4.26), Latitude (11.02→ 10.53), Environment (4.38→3.04)
Late-life history traits			
Flower color[1]	Treatment (*), Sex (***), Anc (*)	−3.76	Treatment (6.21→5.78)
Stem color[1]	Treatment (**), Sex (***), Anc (***)	−1323.61	Longitude (13.5→0.30), Treatment (25.68→12.71)
Dry biomass	Treatment (***), Sex (***), Treat × Sex (***), Lat (*), Anc (.)	−452.90	Latitude (3.00→4.64)

Data were analyzed at the individual level using nested family structure and block effects as random effect covariates. The last column depicts predictors that change in *P*-value class with the addition of ancestry, and their respective shift in their X^2 test statistic.

[1]Phenotypes for which a generalized model was used.

***$P < 0.001$; **$P < 0.01$; *$P < 0.05$; .$P < 0.1$.

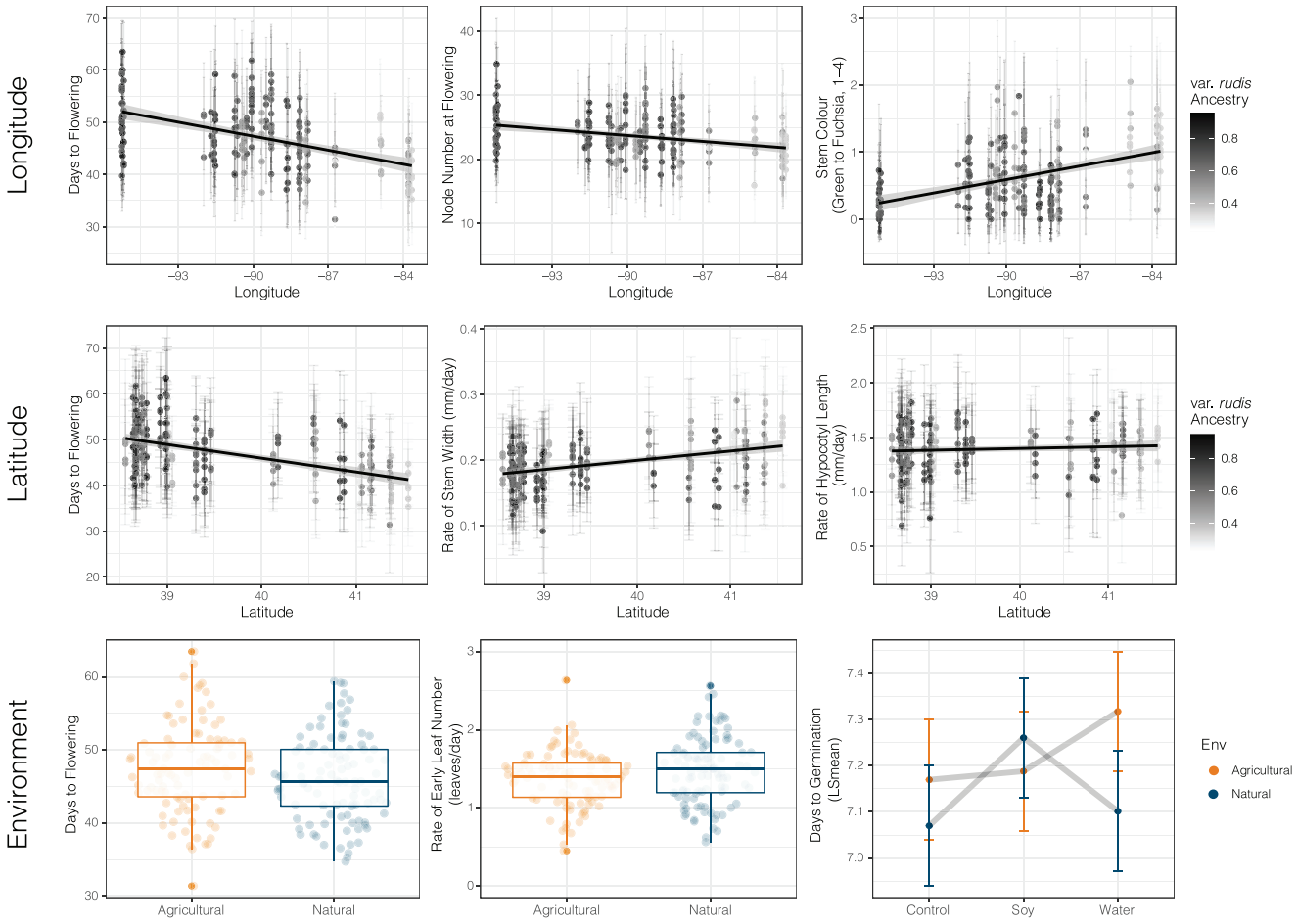

Figure 4. Phenotypic differentiation by longitude, latitude, environment, and the confounding effect of ancestry. All points indicate family-wise means, except for days to germination. (Top) Days to flowering, node number at flowering, and stem color by longitude and proportion of var. *rudis* ancestry. (Middle) Days to flowering, rate of stem width, and rate of hypocotyl length by latitude and proportion of var. *rudis* ancestry. (Bottom) Days to flowering and rate of early leaf number by environment (Left). Least-squares means of days to germination by source environment and treatment, illustrating their interaction (Right).

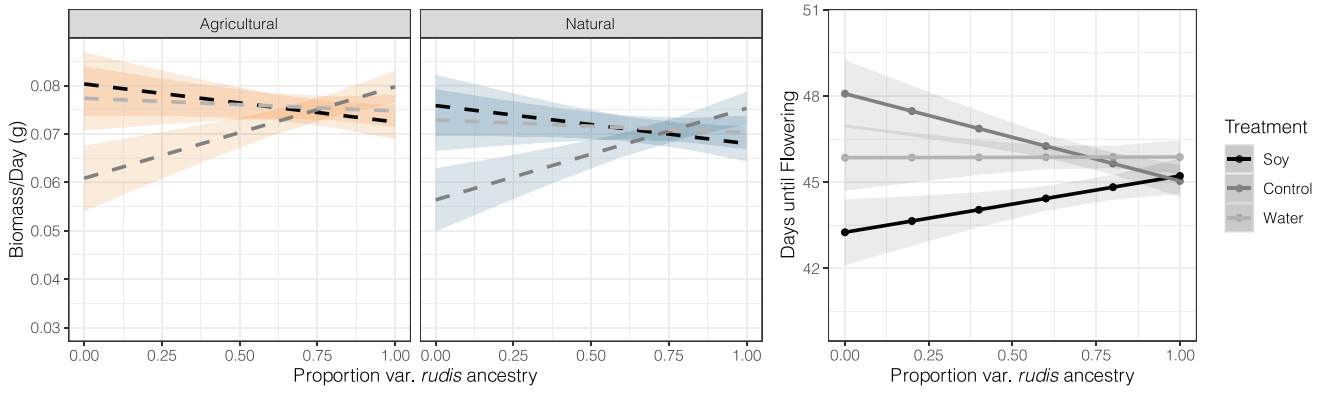

Figure 5. The treatment-dependent effects of var. *rudis* ancestry on male biomass (left panel) and male days to flowering (right). To the illustrate the significant linear, but not quadratic effects of ancestry, the illustrated values are based on the least-squares means of a reduced multiple regression model that excluded the quadratic term, and that controlled for the indirect effects of all other measured phenotypes on fitness. The on average higher biomass of genotypes from agricultural environments (orange) relative to natural environments (blue), regardless of ancestry, is apparent by comparing the two sides of the biomass panel.

increased biomass in the absence of competition and water supplementation (Fig. 5). For male flowering time, of our interest in linear, nonlinear, and interaction effects of ancestry, only the interaction between var. *rudis* ancestry and treatment remained a significant predictor (Ancestry × Treatment: $F_{2,498} = 3.2144$, $P = 0.041015$), with the effect of ancestry on flowering time in soy significantly different from that in the control treatment ($t = 2.514$, $P = 0.0123$). Although higher levels of var. *rudis* ancestry led to a shorter time to flowering in the control treatment, increasing var. *rudis* ancestry extended time to flowering in the soy treatment (Fig. 5). These male, treatment-specific effects of ancestry more broadly reflect a loss of phenotypic plasticity (convergence of time to flowering or rate of biomass, regardless of reared environment) with increasing proportion of var. *rudis* ancestry. Female flowering time and biomass were not influenced by var. *rudis* ancestry (neither linear or quadratic terms, or through its interaction with treatment), but females showed marginally lower biomass and marginally earlier time to flowering in natural compared to agricultural environments (Biomass: $F_{1,498} = 3.091$, $P = 0.0793$; Flowering: $F_{1,498} = 3.374$, $P = 0.0668$).

Discussion

We found evidence of phenotypic differentiation in *A. tuberculatus* across geographic gradients, across ancestral lineages, and resulting from the transition from natural riparian to highly disturbed agricultural environments. Geographic gradients in climate has led to strong latitudinal clines in growth and life history characteristics; however, longitudinal phenotypic clines are in large part explainable by ancestry. Although we found that agricultural populations tended to exhibit longer times to germination in their simulated "away" environment, slower early growth rates, and longer time to flowering, these differences were also in part related to differential ancestry across natural and agricultural environments. We found that the transition of *A. tuberculatus* into agricultural environments has favored southwestern var. *rudis* ancestry—ancestry that leads to lower phenotypic plasticity in fitness-related traits and generally higher biomass, but also treatment-dependent phenotypes. Higher var. *rudis* ancestry results in longer time to flowering in the face of competition, and both faster time to flowering and increased biomass in conditions lacking competition or water supplementation (i.e., the control treatment). When accounting for these complex treatment-dependent effects of ancestry, we also found marginally lower biomass and earlier flowering time in natural compared to agricultural environments, suggesting invasive agricultural populations may be adapting to a new fitness peak. Therefore, phenotypic differentiation among natural and agricultural environments is likely to be driven by altered fitness landscapes that has led to both the selective sorting of var. *rudis* ancestry (preadaptation) and de novo adaptation. These results highlight how human-mediated disturbance and agricultural regimes drive the evolution of native species, shaping interactions between once isolated lineages and drawing from adaptive variation on multiple timescales.

PHENOTYPIC UNDERPINNINGS OF AGRICULTURAL ADAPTATION IN THE FACE OF GENE FLOW

We were interested in the extent to which phenotypic variation consistently differed among natural and agricultural environments in a common garden of highly replicated genotypes from environmentally paired populations across a wide sampling of the *A. tuberculatus* native range. An initial investigation of predictors of individual-level phenotypic variation observed in our common garden experiment, accounting for hierarchical structure of families within populations but before controlling for ancestry, showed that individuals from agricultural populations tend to flower later (1.5 days), have fewer leaves early on in their life history, and suggested local adaptation via germination (through an environment × treatment effect "home advantage") (Fig. 4). Although time to germination was relatively similar for agricultural and natural types in both the control and soy treatment, natural types germinated significantly earlier than agricultural types in the water treatment, suggesting that agricultural adaptation via germination may be driven by moisture availability rather than competition. Interestingly, in large part these phenotypic differences were not consistent with the hypothesis that disturbance regimes will select for accelerated life history (i.e., days to flowering) in agricultural populations (De Wet and Harlan 1975). Overall, although the magnitude of these phenotypic differences between environments appears small, that we observed consistent differences across environments with our paired collections suggests that selection may be acting on these traits despite little observed evolutionary response. The efficacy of the evolutionary response to agriculture may be dampened by very recent timescales of selection, or gene flow hindering an environment-specific response to selection. That we find population pair to be a significant predictor population structure (from both PCAs and Faststructure; Fig. 2) after controlling for latitude and longitude suggests that gene flow in particular may be constraining the response to selection.

A ROLE FOR BOTH PREADAPTATION VIA PREFERENTIAL SORTING OF ANCESTRY AND DE NOVO AGRICULTURAL ADAPTATION

Gene flow across environments and across the range may not only lead to reduced differentiation in phenotypes but may also drive heterogeneity in shared ancestry. In the case of *A. tuberculatus* varieties, secondary contact between these two ancestral

lineages has led to longitudinal clines in ancestry across their range. We find that in large part, the longitudinal clines in phenotypes we initially observed (e.g., days to flowering, stem color) covary with longitudinal clines in ancestry (Figs. 2 and 4), implying that differences that have accumulated in allopatry among *A. tuberculatus* lineages have resulted in phenotypic differentiation in not just seed dehiscence, seedling and flower morphology as has been described (Sauer 1955, Costea et al. 2005), but also in key life history characteristics that appear to be driven by broad-scale adaptation to climate. Stem color shows the most extreme pattern of phenotypic differentiation by ancestry that we observe, with northeastern var. *tuberculatus* ancestry displaying significantly darker purple coloring compared to lighter and greener var. *rudis* stems. This coloration difference among *A. tuberculatus* lineages is consistent with adaptive physiological hypotheses for colder temperatures and northern climates resulting in genetically darker, less reflective coloring (Chalker-Scott 1999; Dick et al. 2011; Koski and Galloway 2020). Compared to longitude, fewer latitudinal clines in growth-related phenotypes were confounded with genetic ancestry, possibly due to the smaller latitudinal variation sampled; however, accounting for ancestry tended to increase latitudinal explanatory power. We found the strongest evidence of latitudinal clines in mid-life history traits—height at flowering, stem width at flowering, and days to flowering (Table 1; Fig. 4)—a signal of broad-scale geographic adaptation to climate with populations evolved in colder climates growing faster and flowering earlier to avoid severe winters (Stinchcombe et al. 2004).

Beyond a longitudinal cline in ancestry, we find that var. *rudis* ancestry is preferentially retained (or var. *tuberculatus* ancestry selected against) in agricultural environments, finding on average 8% higher var. *rudis* ancestry in agricultural environments and up to 44% more within the most extreme natural-agricultural population pairing (Fig. 2). That the agricultural invasion of *A. tuberculatus* has been more severe in southwestern parts of the range may lead one to predict that var. *rudis* ancestry would be associated with agriculture regardless of a role of selection. Furthermore, although past work has shown increased levels of admixture in agricultural compared to natural environments, the extent to which broader-scale processes influenced these patterns remained unclear (Waselkov and Olsen 2014). We explicitly accounted for the two different timescales potentially driving patterns of ancestry—environmental drivers of ancestry on contemporary timescales, and geographic drivers of ancestry on deeper timescales—through sampling pairs of natural and agricultural populations <25 km apart in a replicated fashion across the range. Combining this sampling design with common garden phenotyping and whole-genome sequencing thus provided a powerful test of a key hypothesis put forward in *Evolution* over 60 years ago (Sauer 1957)—the extent to which genetic variation underlying weediness may have predated the association of *A. tuberculatus* with agriculture.

We explicitly tested for agricultural preadaptation (selective sorting of var. *rudis* ancestry) by examining whether the effect of the proportion of var. *rudis* ancestry on phenotypes varied across treatments, which we designed to mimic key components of natural and agricultural environments. We found that the effect of var. *rudis* ancestry on fitness-related traits depended on the reared environment (treatment), implying locally adapted ancestral variation. However, treatment-specific ancestry effects were not necessarily dominated by var. *rudis* types outperforming var. *tuberculatus* types in the soy competition treatment (i.e., a major axis of agricultural habitats) and underperforming in the water treatment (i.e., a major axis of natural habitats), as we predicted. For male biomass, the ancestry by treatment interaction effect was driven by the strong positive effect of var. *rudis* ancestry on biomass in the control treatment, and relative lack thereof in either the soy or water treatments (Fig. 5). Similarly, var. *rudis* ancestry in males led to earlier time to flowering in the control treatment, little difference in the water treatment, and later time to flowering in the soy treatment (Fig. 5). One hypothesis for the potential benefit of later flowering in the presence of focal crops like soy or corn is that it may facilitate a longer vegetative growth period allowing *Amaranthus* to dominate the canopy and facilitate efficient pollen dispersal. However, with drought having been shown to select for early flowering genotypes who shorten their life history in response to a shortened growing season (Cohen 1976; Kozłowski 1992; Franks et al. 2007), our results suggest that the early flowering of var. *rudis* in control treatments—which experienced increased water stress compared to the water treatment—is likely adaptive. Indeed, it is important to note that as opposed to water saturation of the soil, on the hot sunny roof where our experiment was conducted, the water treatment only reduced the severity of soil dry out relative to the control. Thus, the pronounced importance of var. *rudis* ancestry in the control treatment for both flowering- and biomass-based fitness components suggests that var. *rudis* ancestry may experience a selective advantage over var. *tuberculatus* ancestry in drier conditions, as is typical of ruderal habitats, and in agricultural conditions with increasingly frequent droughts. Although these hypotheses need further testing to validate to how these treatment-dependent effects translate in field conditions, the overrepresentation of var. *rudis* ancestry in agricultural environments, higher biomass of pure (but not intermediate) var. *rudis* types, and significant treatment-dependent phenotypic effects of ancestry provides strong evidence for the role of ancestral preadaptation in the *A. tuberculatus* agricultural invasion.

In addition to preadaptation, our investigations suggest that ongoing local adaptation, but not phenotypic plasticity, is further facilitating any selective advantage that var. *rudis* lineages

may have in agronomic environments. We found that regardless of the proportion of var. *rudis* ancestry, natural and agricultural samples showed adaptive differences in germination depending on moisture availability, and that biomass was marginally larger and female flowering time marginally later in individuals from agricultural environments. The evolution of higher fitness in introduced as opposed to native ranges has often been reported (Leger and Rice 2003; Erfmeier and Bruelheide 2005; Caño et al. 2008)—consistent with de novo adaptation to novel agricultural environments facilitating *A. tuberculatus* reaching a new fitness peak. We find no substantial evidence of increased plasticity in genotypes collected from agricultural habitats compared to those from natural habitats, in contrast to weed generalist "jack of all-trades" hypotheses (Richards et al. 2006) that increased plasticity may facilitate the invasion of disturbed agricultural environments. On the contrary, we find that var. *rudis* ancestry, which has been preferentially retained in agricultural environments, shows much less plasticity in both biomass and flowering time, facilitating higher fitness in more diverse environments. This joint inference of the role of ancestry, home environment, reared environment, and geography in shaping patterns of phenotypic variation has thus provided evidence for the invasion of *A. tuberculatus* into agricultural habitats through adaptation across multiple timescales.

CONCLUSIONS

In conclusion, this work has illustrated the power of joint genomic and phenotypic investigation, and the importance of ancestry inference in testing hypotheses about the timescale of adaptation. We find strong evidence for a role of preadaptation in the *A. tuberculatus* invasion of agricultural environments, through the preferential sorting of var. *rudis* ancestry, further supplemented by adaptation on more recent timescales. We show that adaptation to agricultural environments has occurred in the face of gene flow as evidenced by natural-agricultural population proximity predicting similarity of population structure. Future work on the extent of environment-mediated selection for or against gene flow across the genome, the genomic architecture of phenotypic trait differences, and the timescale of allele frequency change associated with agricultural environments will further resolve the enigma of rapid adaptation to human-mediated environmental change.

AUTHOR CONTRIBUTIONS

JMK performed the collections. JMK, SIW, and JRS designed the experiment. JMK and AC performed the common garden experiment. JMK analyzed the data. JMK wrote the first draft. JMK, SIW, and JRS revised the manuscript.

ACKNOWLEDGMENTS

The authors thank J. Santangelo, L. Rieseberg, and the Wright and Stinchcombe labs at the University of Toronto for feedback on the manuscript. The authors tremendously appreciate the assistance of the EEB greenhouse staff—T. Gludovacz and B. Cole—at the University of Toronto for the support throughout the experiment, and assistance from many enthusiastic and hardworking undergraduate researchers without which the common garden experiment could not be performed, especially T. Trepanier, P. Pietraszkiewicz, X. Ma, E. Gajic, T. Lim, and J. To. The authors specially thank C. Morse at the McGregor Herbarium, Kansas University for assisting with collections and providing most westward samples. Lastly, the authors thank three anonymous reviewers and the associate editor for useful feedback and suggestions on the manuscript. Funding for parts of this study was provided by the Society for the Study of Evolutionary Biology Rosemary Grant Advanced Award (JMK), NSERC PGS-D (JMK), Discovery Grants from NSERC Canada (SIW and JRS), and SIW's Canada Research Chair in Population Genomics.

DATA ARCHIVING

Read sequence data are available for download at SRA under the BioProject accession number PRJNA770655. Scripts for all analyses and phenotypic datasets are available at https://github.com/jkreinz/Evolution2021-Preadapt.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

LITERATURE CITED

Arnaud, J. -F., S. Fénart, M. Cordellier, and J. Cuguen. 2010. Populations of weedy crop-wild hybrid beets show contrasting variation in mating system and population genetic structure. Evol. Appl. 3:305–318.

Baker, H. G. 1974. The evolution of weeds. Ann. Rev. Ecol. Syst. 5:1–24.

Barrett, S. C. H. 1982. Genetic variation in weeds. Pp. 73–98 *in* R. C. Charudattan and H. Lynn Walker, eds. Biological control of weeds with plant pathogens. John Wiley, New York.

Barrett, S. H. 1983. Crop mimicry in weeds. Econ. Bot. 37:255–282.

Bilinski, P., P. S. Albert, J. J. Berg, J. A. Birchler, M. N. Grote, A. Lorant, J. Quezada, K. Swarts, J. Yang, and J. Ross-Ibarra. 2018. Parallel altitudinal clines reveal trends in adaptive evolution of genome size in Zea mays. PLos Genet. 14:e1007162.

Boudry, P., M. Mörchen, P. Saumitou-Laprade, P. Vernet, and H. Van Dijk. 1993. The origin and evolution of weed beets: consequences for the breeding and release of herbicide-resistant transgenic sugar beets. Theor. Appl. Genet. 87:471-478.

Broad-Institute. 2016. Picard tools.

Brown, A. H. D., and D. R. Marshall. 1981. Evolutionary changes accompanying colonization in plants. Pp. 351–363 *in* G. G. E. Scudder and J. L. Reveal, eds. Proceedings Second International Congress and Evolutionary Biology. Hunt Institute, Pittsburgh, PA.

Caño, L., J. Escarré, I. Fleck, J. M. Blanco-Moreno, and F. X. Sans. 2008. Increased fitness and plasticity of an invasive species in its introduced range: a study using Senecio pterophorus: increased plasticity of an invasive species. J. Ecol. 96:468–476.

Chalker-Scott, L. 1999. Environmental significance of anthocyanins in plant stress responses. Photochem. Photobiol. 70:1–9.

Charbonneau, A., D. Tack, A. Lale, J. Goldston, M. Caple, E. Conner, O. Barazani, J. Ziffer-Berger, I. Dworkin, and J. K. Conner. 2018. Weed evolution: genetic differentiation among wild, weedy, and crop radish. Evol. Appl. 11:1964–1974.

Clements, D. R., and A. Ditommaso. 2011. Climate change and weed adaptation: can evolution of invasive plants lead to greater range expansion than forecasted? Weed Res. 51:227–240.

Cohen, D. 1976. The optimal timing of reproduction. Am. Nat. 110:801–807.

Cornille, A., A. Salcedo, H. Huang, D. Kryvokhyzha, K. Holm, X.-J. Ge, J. R. Stinchcombe, S. Glémin, S. I. Wright, and M. Lascoux. 2018. Local

adaptation and maladaptation during the worldwide range expansion of a selffertilizing plant. bioRxiv https://doi.org/10.1101/308619.

Costea, M., and F. J. Tardif. 2003. Conspectus and notes on the genus *Amaranthus* in Canada. Rhodora 105:260–281.

Costea, M., S. E. Weaver, and F. J. Tardif. 2005. The biology of invasive alien plants in Canada. 3. *Amaranthus tuberculatus* (Moq.) Sauer var. *rudis* (Sauer) Costea & Tardif. Can. J. Plant Sci. 85:507–522.

De Wet, J. M. J., and J. R. Harlan. 1975. Weeds and domesticates: evolution in the man-made habitat. Econ. Bot. 29:99–107.

Dick, C. A., J. Buenrostro, T. Butler, M. L. Carlson, D. J. Kliebenstein, and J. B. Whittall. 2011. Arctic mustard flower color polymorphism controlled by petal-specific downregulation at the threshold of the anthocyanin biosynthetic pathway. PLoS One 6:e18230.

Erfmeier, A., and H. Bruelheide. 2005. Invasive and native *Rhododendron ponticum* populations: is there evidence for genotypic differences in germination and growth? Ecography 28:417–428.

Exposito-Alonso, M. 2020. Seasonal timing adaptation across the geographic range of *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. U S A. 117:9665–9667.

Fang, H. 2014. Towards better understanding of artifacts in variant calling from high-coverage samples. Bioinformatics 30:2843–2851.

Fisher, R. A. 1930. The genetical theory of natural selection: a complete variorum edition. Oxford Univ. Press, Oxford, U.K.

Frachon, L., C. Bartoli, S. Carrère, O. Bouchez, A. Chaubet, M. Gautier, D. Roby, and F. Roux. 2018. A genomic map of climate adaptation in arabidopsis thaliana at a micro-geographic scale. Front. Plant Sci. 9:967.

Franks, S. J., S. Sim, and A. E. Weis. 2007. Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. Proc. Natl. Acad. Sci. U S A. 104:1278–1282.

Garrison, E., and G. Marth. 2012. Haplotype-based variant detection from short-read sequencing. arXiv. 1207.3907.

Guo, W. Y., C. Lambertini, L. X. Nguyen, X. Z. Li, and H. Brix. 2014. Preadaptation and post-introduction evolution facilitate the invasion of *Phragmites australis* in North America. Ecol. Evol. 4:4567-4577.

Holm, K. 2010. Studies on natural variation and evolution of photoperiodism in plants. Doctoral thesis. Uppsala University, Uppsala, Sweden.

Kane, N. C., and L. H. Rieseberg. 2008. Genetics and evolution of weedy *Helianthus annuus* populations: adaptation of an agricultural weed. Mol. Ecol. 17:384–394.

Koski, M. H., and L. F. Galloway. 2020. Geographic variation in floral color and reflectance correlates with temperature and colonization history. Front. Plant Sci. 11:991.

Kozłowski, J. 1992. Optimal allocation of resources to growth and reproduction: implications for age and size at maturity. Trends Ecol. Evol. 7:15–19.

Kreiner, J. M., D. A. Giacomini, F. Bemm, B. Waithaka, J. Regalado, C. Lanz, J. Hildebrandt, P. H. Sikkema, P. J. Tranel, D. Weigel, et al. 2019. Multiple modes of convergent adaptation in the spread of glyphosate-resistant *Amaranthus tuberculatus*. Proc. Natl. Acad. Sci. U S A. 116:21076–21084.

Kreiner, J. M., G. Sandler, A. J. Stern, P. J. Tranel, D. Weigel, J. R. Stinchcombe, and S. I. Wright. 2021. Repeated origins, gene flow, and allelic interactions of herbicide resistance mutations in a widespread agricultural weed. bioRxiv https://doi.org/10.1101/2021.05.10.443516.

Kuester, A., J. K. Conner, T. Culley, and R. S. Baucom. 2014. How weeds emerge: a taxonomic and trait-based examination using United States data. New Phytol. 202:1055-1068.

Lai, Z., N. C. Kane, A. Kozik, K. A. Hodgins, K. M. Dlugosch, M. S. Barker, M. Matvienko, Q. Yu, K. G. Turner, S. A. Pearl, et al. 2012. Genomics of Compositae weeds: EST libraries, microarrays, and evidence of introgression. Am. J. Bot. 99:209–218.

Leger, E. A., and K. J. Rice. 2003. Invasive California poppies (*Eschscholzia californica* Cham.) grow larger than native individuals under reduced competition. Ecol. Lett. 6:257–264.

Leon, R. G., D. C. Bassham, and M. D. K. Owen. 2007. Thermal and hormonal regulation of the dormancy? Germination transition in *Amaranthus tuberculatus* seeds. Weed Res. 47:335–344.

Leon, R. G., A. D. Knapp, and M. D. K. Owen. 2004. Effect of temperature on the germination of common waterhemp (*Amaranthus tuberculatus*), giant foxtail (*Setaria faberi*), and velvetleaf (*Abutilon theophrasti*). Weed Sci. 52:67–73.

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2 [q-bio.GN].

Liebman, M., C. L. Mohler, and C. P. Staver. 2001. Ecological management of agricultural weeds. Cambridge Univ. Press, Cambridge, U.K.

Lotterhos, K. E., and M. C. Whitlock. 2015. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Mol. Ecol. 24:1031–1046.

Martin, S. L., J.-S. Parent, M. Laforest, E. Page, J. M. Kreiner, and T. James. 2019. Population genomic approaches for weed science. Plants 8:354.

McGoey, B. V., R. Janik, and J. R. Stinchcombe. 2017. Individual chambers for controlling crosses in wind-pollinated plants. Methods Ecol. Evol. /British Ecological Society 8:887–891.

Muller, M.-H., M. Latreille, and C. Tollon. 2011. The origin and evolution of a recent agricultural weed: population genetic diversity of weedy populations of sunflower (*Helianthus annuus* L.) in Spain and France. Evol. Appl. 4:499–514.

Murphy, C. E., and D. Lemerle. 2006. Continuous cropping systems and weed selection. Euphytica/Netherlands journal of plant breeding 148:61–73.

Neuffer, B. 1990. Ecotype differentiation in Capsella. Vegetatio 89:165–171.

Peterson, M. L., K. M. Kay, and A. L. Angert. 2016. The scale of local adaptation in *Mimulus guttatus*: comparing life history races, ecotypes, and populations. New Phytol. 211:345–356.

Pratt, D. B., and C. LG, 2001. *Amaranthus rudis* and *A. tuberculatus*, one species or two? J. Torrey Bot. Soc. 128:282-296.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–575.

Raj, A., M. Stephens, and J. K. Pritchard. 2014. fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics 197:573–589.

Ranallo-Benavidez TR, J. K. S., and M. C. Schatz. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat. Commun. 11:1432.

Ravet, K., E. L. Patterson, H. Krähmer, K. Hamouzová, L. Fan, M. Jasieniuk, A. Lawton-Rauh, J. M. Malone, J. S. McElroy, A. Merotto Jr., et al. 2018. The power and potential of genomics in weed biology and management. Pest Manage. Sci. 74:2216–2225.

Richards, C. L., O. Bossdorf, N. Z. Muth, J. Gurevitch, and M. Pigliucci. 2006. Jack of all trades, master of some? On the role of phenotypic plasticity in plant invasions. Ecol. Lett. 9:981–993.

Riddell, J. L. 1835. A synopsis of the flora of the western states. E Deming, Cincinnati, OH.

Hartzler, R. G., B. A. Battles, and D. Nordby. 2004. Effect of common waterhemp (*Amaranthus rudis*) emergence date on growth and fecundity in soybean. Weed Sci. 52:242–245.

Samis, K. E., K. D. Heath, and J. R. Stinchcombe. 2008. Discordant longitudinal clines in flowering time and phytochrome C in *Arabidopsis thaliana*. Evolution 62:2971–2983.

Samis, K. E., C. J. Murren, O. Bossdorf, K. Donohue, C. B. Fenster, R. L. Malmberg, M. D. Purugganan, and J. R. Stinchcombe. 2012. Longitudinal trends in climate drive flowering time clines in North American *Arabidopsis thaliana*. Ecol. Evol. 2:1162–1180.

Sauer, J. 1955. Revision of the dioecious amaranths. Madroño 13:5–46.

Sauer, J. 1957. Recent migration and evolution of the dioecious amaranths. Evolution 11:11–31.

Sellers, B. A., R. J. Smeda, W. G. Johnson, J. Andrew Kendig, and M. R. Ellersieck. 2003. Comparative growth of six Amaranthus species in Missouri. Weed Sci. 51:329–333.

Schlaepfer, D. R., M. Glättli, M. Fischer, and M. van Kleunen. 2010. A multi-species experiment in their native range indicates pre-adaptation of invasive alien plant species. New Phytol. 185:1087-1099.

Stevens, O. A. 1932. The number and weight of seeds produced by weeds. Am. J. Bot. 19:784–794.

Stewart, C. N., P. J. Tranel, D. P. Horvath, J. V. Anderson, L. H. Rieseberg, J. H. Westwood, M. -S. CA, M. L. Zapiola, and K. M. Dlugosch. 2009. Evolution of weediness and invasiveness: charting the course for weed genomics. Weed Sci. 57:451–462.

Stinchcombe, J. R., C. Weinig, M. Ungerer, K. M. Olsen, C. Mays, S. S. Halldorsdottir, M. D. Purugganan, and J. Schmitt. 2004. A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene FRIGIDA. Proc. Natl. Acad. Sci. U S A. 101:4712–4717.

Turesson, G. 1992. The species and the variety as ecological units. Hereditas, 3:100–113.

Uline, E. B., and W. L. Bray. 1895. Synopsis of North American Amaranthaceae. III. Bot. gaz. 20:337–344.

Vigueira, C. C., K. M. Olsen, and A. L. Caicedo. 2013. The red queen in the corn: agricultural weeds as models of rapid adaptive evolution. Heredity 110:303–311.

Waselkov, K. E., and K. M. Olsen. 2014. Population genetics and origin of the native North American agricultural weed waterhemp (*Amaranthus tuberculatus*; Amaranthaceae). Am. J. Bot. 101:1726–1736.

Waselkov, K. E., N. D. Regenold, R. C. Lum, and K. M. Olsen. 2020. Agricultural adaptation in the native North American weed waterhemp, *Amaranthus tuberculatus* (Amaranthaceae). PLoS One 15:e0238861.

Ye, C. Y., W. Tang, D. Wu, L. Jia, J. Qiu, M. Chen, L. Mao, F. Lin, H. Xu, X. Yu, et al. 2019. Genomic evidence of human selection on Vavilovian mimicry. Nat. Ecol. Evol. 3:1474-1482.