Correction

EVOLUTION

Correction for "Multiple modes of convergent adaptation in the spread of glyphosate-resistant *Amaranthus tuberculatus*," by Julia M. Kreiner, Darci Ann Giacomini, Felix Bemm, Bridgit Waithaka, Julian Regalado, Christa Lanz, Julia Hildebrandt, Peter H. Sikkema, Patrick J. Tranel, Detlef Weigel, John R. Stinchcombe, and Stephen I. Wright, which was first published September 30, 2019; 10.1073/pnas.1900870116 (*Proc. Natl. Acad. Sci. U.S.A.* **116**, 21076–21084).

The authors note that Fig. 5 and its corresponding legend appeared incorrectly. The corrected figure and its corrected legend appear below. Both versions of the reference genome and their annotations are publicly available on the CoGe database: the pseudoscaffolded reference genome (finished to *Amaranthus hypochondriacus*, 16 pseudochromosomes) under ID 54057, and the original genome (2,514 scaffolds) under ID 51756.

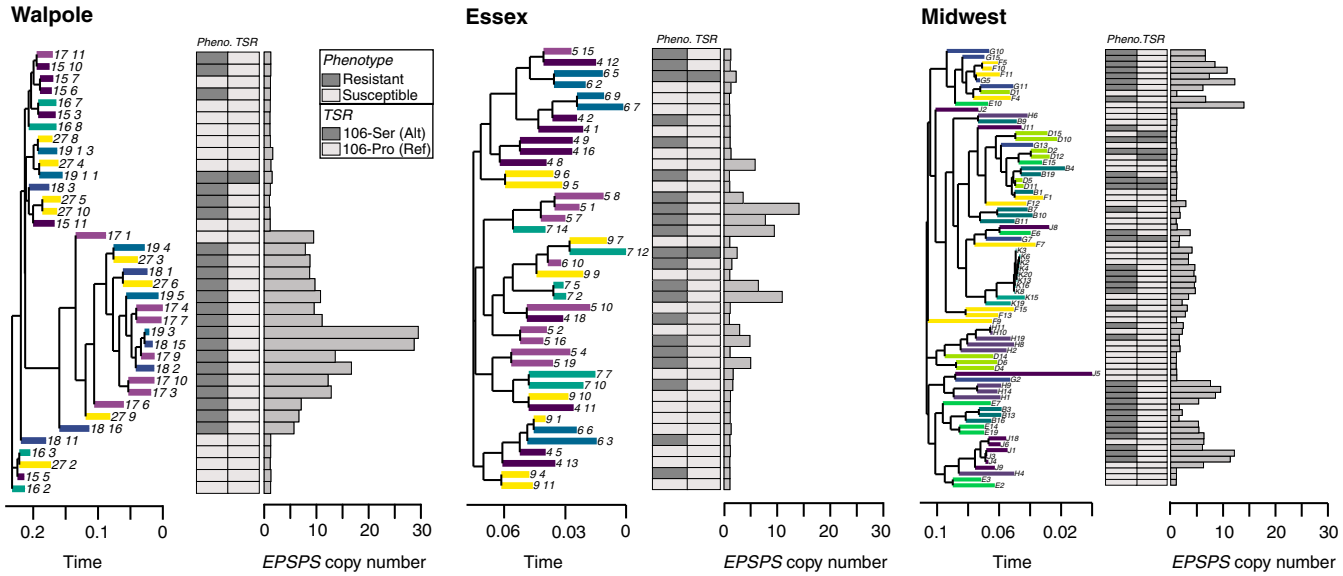

Fig. 5. Diversity of *EPSPS* amplification origins across agricultural regions. For each agricultural region, we show a haplotype tree of based on SNPs within *EPSPS* (with tips colored by population of origin), alongside a bar plot of *EPSPS* copy number, and a matrix of phenotypic resistance and target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

First published November 4, 2019.

www.pnas.org/cgi/doi/10.1073/pnas.1917644116

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

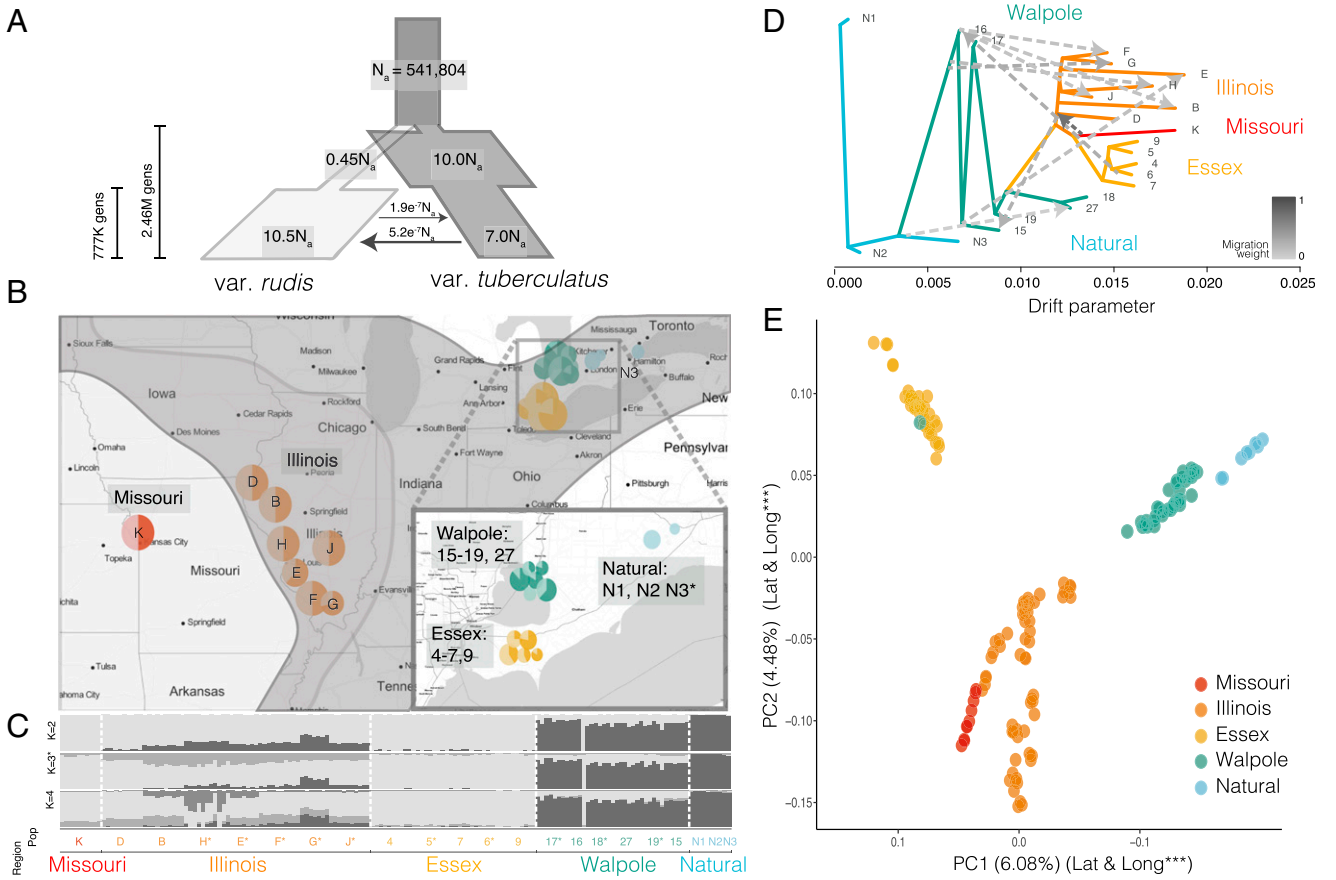

Fig. 1. Population structure and demographic history in *Amaranthus tuberculatus*. (*A*) Demographic model of the 2 *A. tuberculatus* subspecies (N_a = ancestral effective population size, with migration and other effective population size estimates scaled accordingly). (*B*) Geographic map of phenotyped 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

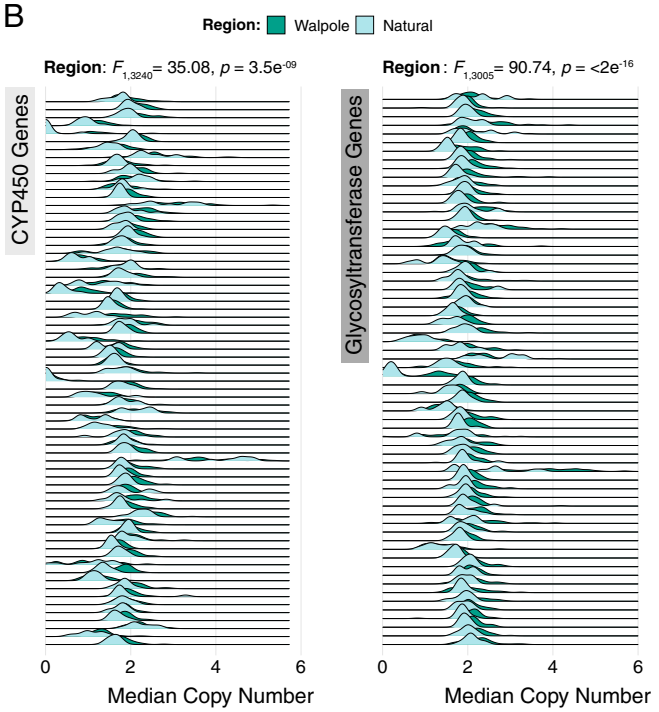

Fig. 2. Enrichment and expansion of CYP450 and glycosyltransferase gene families in the transition from natural to agricultural in Walpole. (*A*) GO categories that were significantly enriched in an analysis of the 99th percentile of F_{st} outliers between Walpole and Natural populations. Light gray 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 peptidyl-amino 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

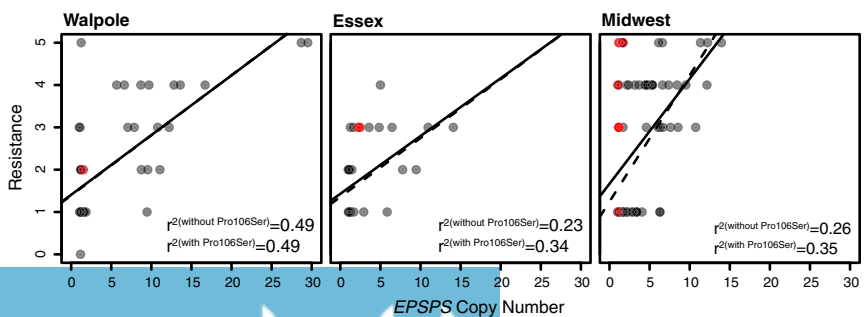

Fig. 3. *EPSPS* copy number variation among individuals, and its relationship with resistance. *EPSPS* copy number significantly explains phenotypic resistance within each agricultural region (solid linear regression line) (Walpole, P = 2.6e-07; Essex, P = 0.002; Midwest, P = 3.5e-06), with P106S substitution 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

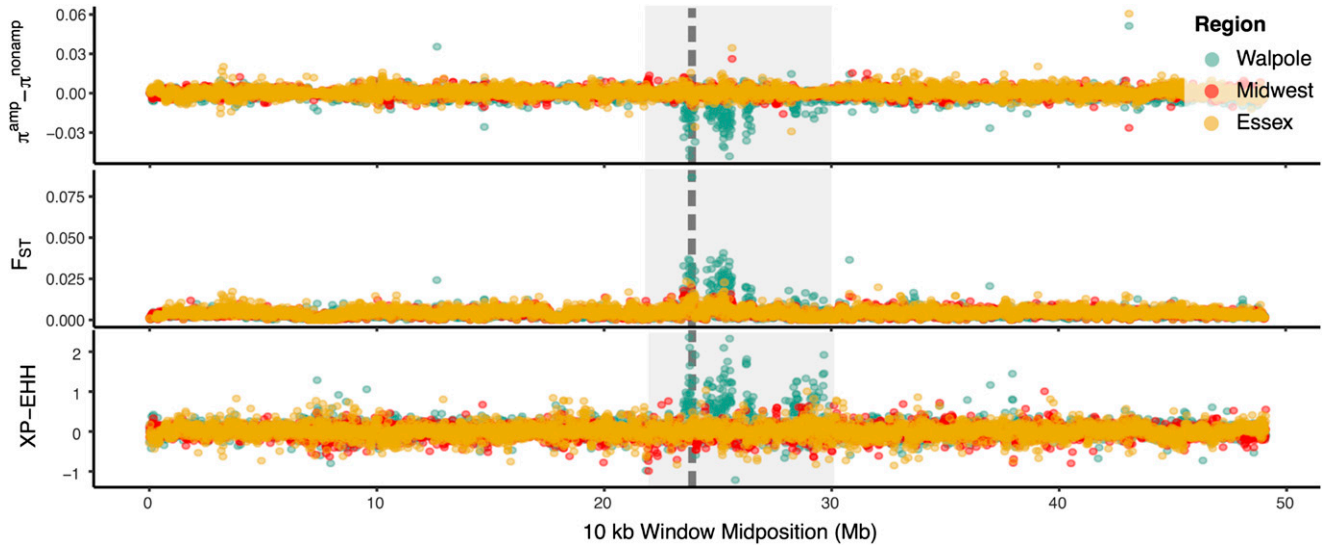

Fig. 4. Population genetic signals of selection related to copy number increase around *EPSPS* on chromosome 5. The deficit of diversity (*Top*), relative differentiation (*Middle*), and difference in extended haplotype homozygosity [XP-EHH (34)] (*Bottom*) is compared between amplified and nonamplified 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.

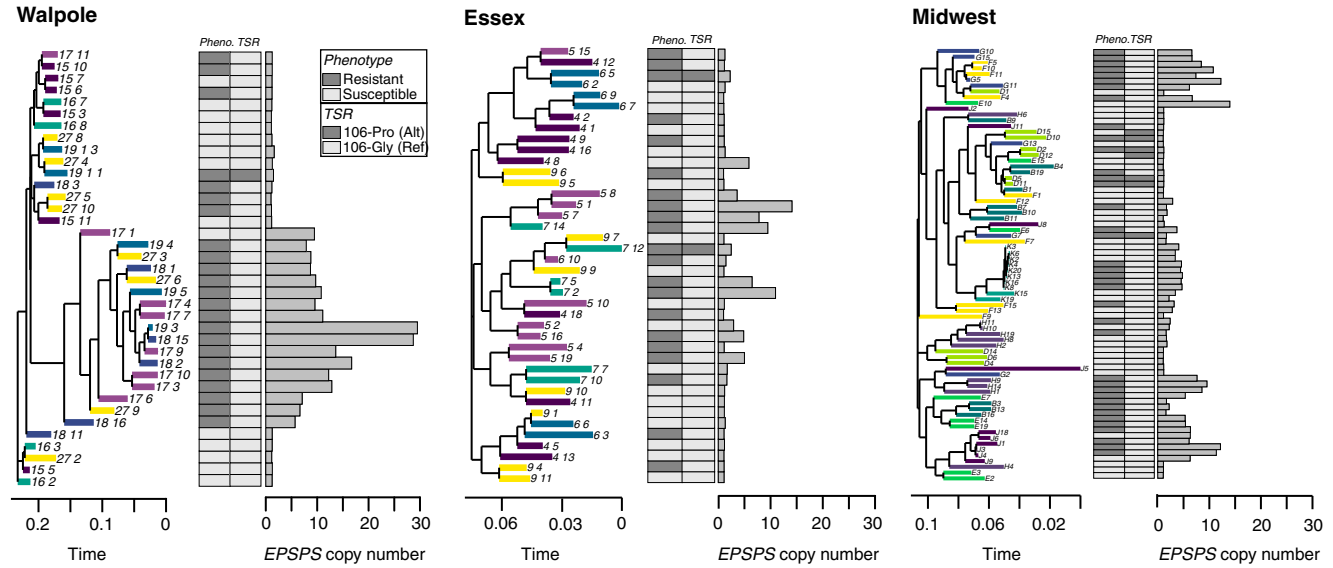

Fig. 5. Diversity of *EPSPS* amplification origins across agricultural regions. For each agricultural region, we show a haplotype tree of based on SNPs within *EPSPS* (with tips colored by population of origin), alongside a bar plot of *EPSPS* copy number, and a matrix of phenotypic resistance and target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

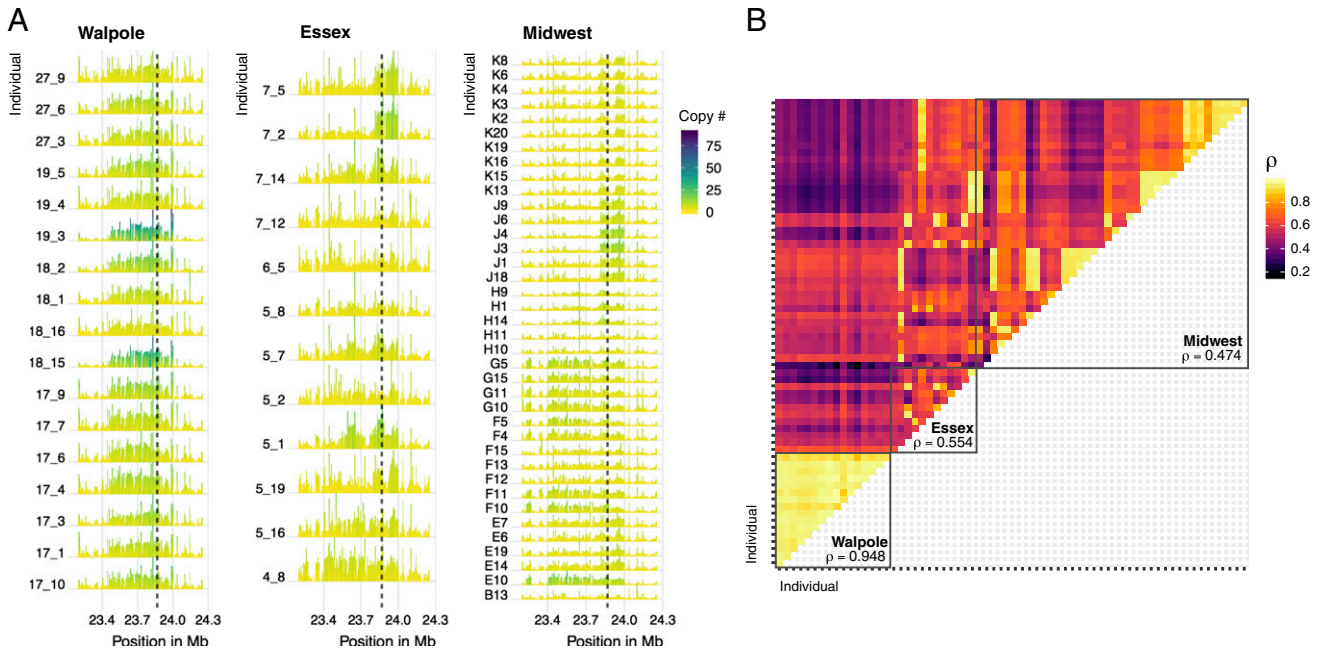

Fig. 6. Similarity of the *EPSPS* amplification inferred from copy number variation around the *EPSPS* gene. (*A*) The profile of normalized sequence coverage for 1 Mb around EPSPS (locus delimited by dashed line). Color of each bar under the curve represents copy number at that genomic position, where 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).