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

Julia M. Kreiner[1,*], Darci Giacomini[2], Felix Bemm[3,5], Bridgit Waithaka[3], Julian Regalado[3], Christa Lanz[3], Julia Hildebrandt[3], Peter H. Sikkema[4], Patrick J. Tranel[2], Detlef Weigel[3,*], John R. Stinchcombe[1], and Stephen I. Wright[1]

[1]Department of Ecology Evolutionary Biology, University of Toronto, Toronto, ON, Canada
[2]Department of Crop Sciences, University of Illinois, Urbana, IL, USA
[3]Department of Molecular Biology, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany
[4]Department of Plant Agriculture, University of Guelph Ridgetown Campus, Ridgetown, ON N0P 2C0, Canada.
[5]Current address: KWS SE, Grimsehlstraße 31, 37555 Einbeck, Germany

Correspondence: ***julia.kreiner@mail.utoronto.ca, weigel@tue.mpg.de***

convergent adaptation | evolution | herbicide resistance | independent adaptation | gene flow | standing variation | Amaranthus

The selection pressure exerted by herbicides has led to the repeated evolution of resistance in weeds. The evolution of herbicide resistance on contemporary timescales provides an outstanding opportunity to investigate key open questions about the genetics of adaptation, in particular the relative importance of adaptation from new mutations, standing genetic variation, and 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 (1), with both agricultural populations and resistance only recently emerging in Canada (2, 3). 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 in Canada and the USA. 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 non-agricultural. Moreover, resistance on these local, non-agricultural backgrounds appears to have occurred predominantly through the partial sweep of a single amplification haplotype. In contrast, US genotypes and those in Canada introduced from the US show multiple amplification haplotypes segregating both between and within populations. Therefore, while the remarkable diversity of *A. tuberculatus* has facilitated geographic parallel adaptation of glyphosate resistance, different timescales of selection have favored either adaptation from standing variation or *de novo* mutation in certain parts of the range.

Glyphosate-resistant *A. tuberculatus* was first reported in Missouri in 2005, but has since been documented in 19 American states (1), with resistant biotypes harming corn and soybean yields (3, 4). Agriculturally-associated *A. tuberculatus* emerged in Canada in the province of Ontario only in the early 2000's, with reports of glyphosate resistance following a decade later (2, 3). As with other herbicides, resistance 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 (5–9). More often, glyphosate resistance in *Amaranthus* has an unusual genetic basis: amplification of the *EPSPS* locus (10–14). Gene amplification apparently evolved independently in two *Amaranthus* species (13–16), raising the possibility that it could have evolved multiple times independently (17). While glyphosate resistance has been studied from multiple angles (18–22), the recent discovery of glyphosate-resistant *A. tuberculatus* in southwestern Ontario affords the unique opportunity to evaluate evolutionary origins and processes driving the recent spread of herbicide resistance in an agronomically important weed.

We assembled a high-quality reference genome for *A. tuberculatus* from a single individual with 58 Gb (approx. 87X genome coverage) of 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 Sup Table 1 for details). Our final genome size is consistent with recent cytometric estimates of 676 Mb (SE=27 Mb) for *A. tuberculatus* (23). The new reference included 88% of the near-universal single copy orthologs present in BUSCO's Embryophyta benchmarking dataset with 6% marked as duplicate (24). For chromosome-scale sweep scan analyses, we further scaffolded our contigs onto the resolved *A. hypochondriacus* genome (25), resulting in 16 final pseudomolecules for analysis, including 99.8% of our original assembly (see methods).

We resequenced whole genomes of 163 individuals to 10X coverage from 19 agricultural fields in Missouri, Illinois, and two regions where glyphosate resistance has recently appeared in Ontario—Essex County, an agriculturally important region in southwestern Ontario, and Walpole Island, an expansive wetland with growing agricultural activity. We also sampled 10 individuals from natural populations in Ontario as a non-agricultural, native Canadian comparison. Genome-wide diversity in *A. tuberculatus* is extremely high, even relative to other wind-pollinated outcrossers (26), with diversity at four-fold degenerate sites = 0.041. The frequencies of glyphosate resistance in our focal agricultural fields ranged from 13% to 88%, based on greenhouse trials (see

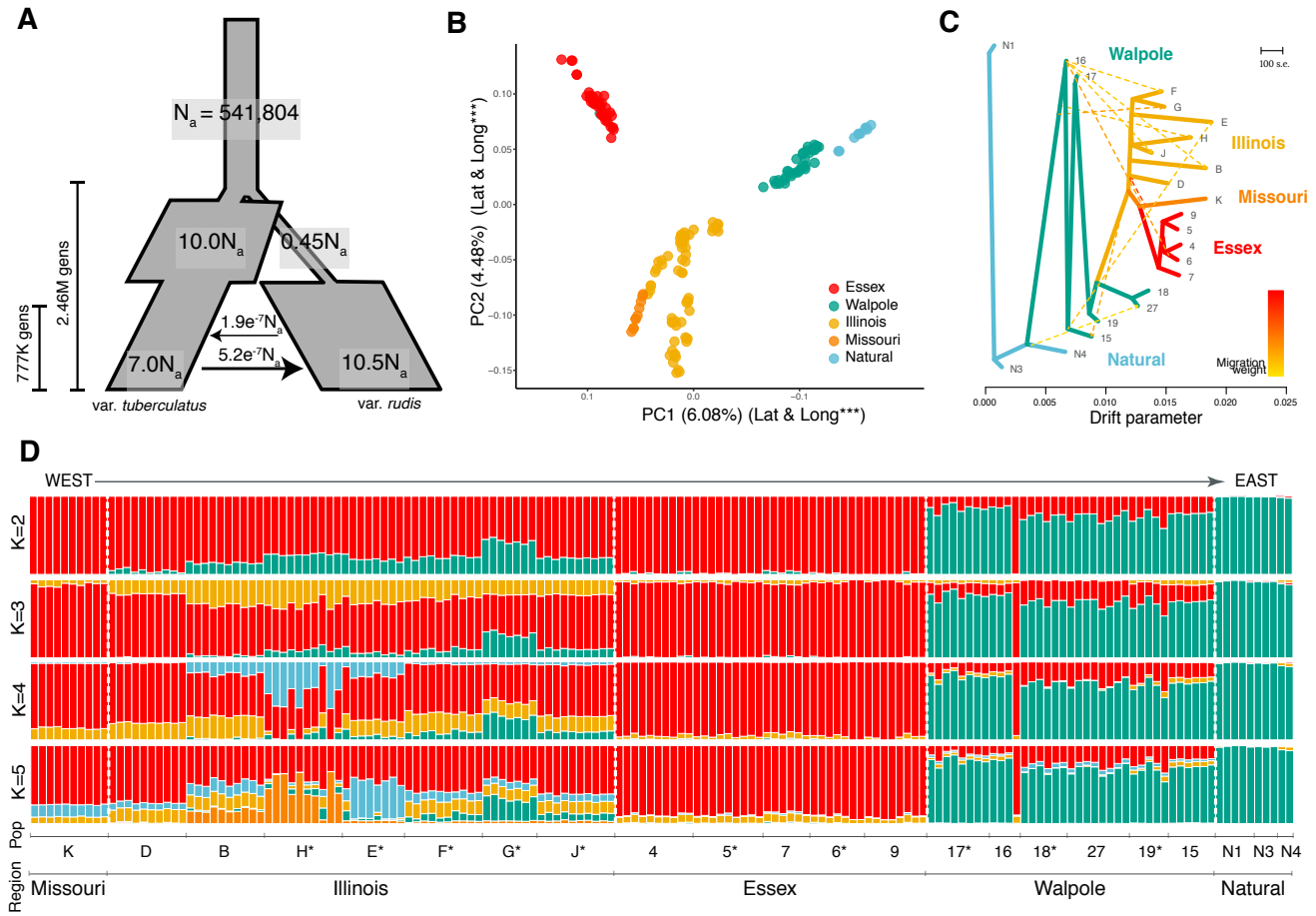


Fig. 1. Population structure and demographic history in *A. tuberculatus*. A) Demographic model of *A. tuberculatus* subspecies. N_a refers to ancestral effective population size. B) A PCA of genotypes from all samples across Ontario and the Midwest, with both PC1 and PC2 significantly relating to both longitude and latitude. C) *Treemix* results showing the maximum likelihood of relatedness between all populations based on allele frequencies. D) *STRUCTURE* plot of admixture across regions and populations from west (left) to east (right), with most likely number of clusters (K) = 2. Predominantly resistant populations are indicated by an asterix.

Methods). Samples from natural populations in Ontario had no glyphosate resistance.

To dissect the genetic origins of convergent adaptation to glyphosate across the sampled range, we first characterized genome-wide patterns of population structure, demography, and differentiation. Population structure, demographic modelling (Fig 1), and phenotypic characterization confirmed the presence of two previously hypothesized ancestral lineages (27, 28): *A. tuberculatus* var. *rudis*, which arose in the southern midwest US and is thought to be pre-adapted to agricultural environments (27, 28), and *A. tuberculatus* var. *tuberculatus*, a variety native to the northeast US and Canada, found primarily in riparian environments (3). Population structure largely reflects historical range limits (28): natural Ontario populations possess the diagnostic indehiscent seed phenotype and are genetically homogeneous for ancestry of the var. *tuberculatus* lineage, Missouri samples are homogeneous for the var. *rudis* lineage, while Illinois, a region of sympatry in the historical range of the two subspecies, is admixed, with the amount of admixture generally increasing from west to east (Fig 1C,D). The most likely tuberculatus-rudis demographic model is one of secondary contact, with var. *rudis* having undergone a bottleneck followed by a dramatic expansion (Fig 1A). Therefore, population genomic analyses largely support the interpretation of two varieties diverging on either side of the Mississippi river, which were recently brought back into contact through human-mediated expansion of var. *rudis*.

Analysis of agricultural populations in Ontario, which have only recently become problematic, shed new light on the demographic source of the *A. tuberculatus* invasion. Populations from Essex county fall completely within the var. *rudis* cluster, with a *treemix* model indicating that Essex populations are most closely related to the most western Missouri population (Fig 1C), from which 99.6% of the Essex genome derived (*f* statistic, (29)). These patterns of population structure are distinct from the continuous gradient of west-east ancestry previously reported (27), and support the hypothesis that glyphosate-resistant *A. tuberculatus* was introduced to Ontario through seed-contaminated agricultural machinery (3) or animal-mediated seed dispersal (30).

In contrast, populations from Walpole Island, where glyphosate resistance was first reported in Ontario (2), are mainly of the eastern var. *tuberculatus* type (Fig 1). Con-

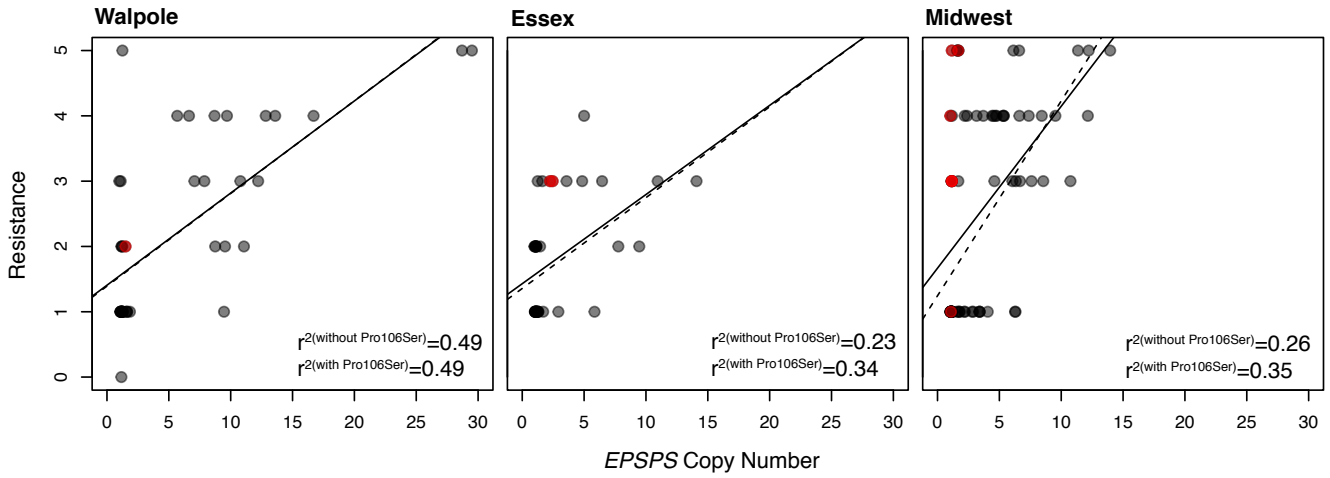


Fig. 2. *EPSPS* copy number variation among individuals, and its relationship with resistance. A) *EPSPS* copy number significantly explains phenotypic resistance within each agricultural region (solid linear regression line), with proline-106-serine *EPSPS* substitution (in red) substantially increasing the variation explained in Essex and the Midwest, but not in Walpole (dashed linear regression line).

sistent with structure analyses, Walpole populations are the least differentiated from nearby natural populations (D_{XY} = 0.0448; Sup Fig 1). This is surprising given past suggestions that var. *rudis* ancestry is a key prerequisite for agricultural invasion (3, 27), and suggests that these populations may have experienced strong and rapid local adaptation upon the conversion of wetlands to agricultural fields. A GO enrichment test for the top 1% of loci with excess differentiation between Walpole and natural populations for a given level of Walpole diversity (representing loci that have undergone positive selection in Walpole; Sup Fig 2) showed significant enriched for biological processes involved in gene expression, RNA processing, and several metabolism related classes (Sup Table 3) that broadly function in flowering time, growth, stress response, nitrogen metabolism, and heavy metal detoxification. Flowering time and growth rate have previously been identified as playing a role in *A. tuberculatus* agricultural adaptation (31), while the other traits are strong candidates for future investigations of phenotypes and genes that underlie high fitness in these frequently disturbed environments that experience a wide array of synthetic inputs. Moreover, mutations in one outlier gene identified, *CHY1*, a peroxisomal hydrolase involved in cold tolerance (32), has been shown in Arabidopsis to confer resistance to the herbicide 2,4-dichlorophenoxybutyric acid (2,4-DB) by preventing its $\beta-$oxidation to toxic 2,4-D(33).

While Walpole presents itself as a striking example of rapid agricultural adaptation, whereas Essex is an introduction of a preadapted genotype to a new locale, this convergent evolution on agricultural fields may not be solely the result of *de novo* mutations. Populations from Walpole Island show some level of introgression from var. *rudis* (f = 17.8%), while treemix estimates that 9/10 migration events (explaining 2.5% of SNP variation) across all samples involve Walpole (Fig 1C). 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 adaptation to agricultural environments, and possibly the evolution of glyphosate resistance.

Two major evolutionary paths to glyphosate resistance are amplification of wild-type *EPSPS* or non-synonymous mutations 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 two classes of *EPSPS* mutations. Using our genomic data to quantify sequence copy number (see methods), we found that of 84 individuals assayed in the greenhouse as resistant, 60 (71%) had elevated *EPSPS* copy number. While *EPSPS* amplification was most frequent in the Midwest (82.5% of resistant individuals, compared to 70% in Walpole and 52% in Essex), the magnitude of the amplification in resistant individuals was on average almost twice in Walpole (~9 copies, compared to 5 in the Midwest, and 4 in Essex). Previous estimates of *EPSPS* copy number in resistant *A. tuberculatus* are up to 17.5X that of diploid susceptibles (10); we found two individuals in Walpole with 29X copy number (Fig 2). A regression of resistance onto copy number was significant in all three 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, where an additional 10% of variation was explained by a non-synonymous change, proline-106-serine (Fig 2).

Our chromosome-scale genome assembly provided a unique opportunity to determine the genomic footprint of selection around *EPSPS* in different populations. Across all populations, the *EPSPS* amplification was much more extensive than the 10 kb *EPSPS* gene—phenotypically resistant individuals were characterized by an increase in copy number mean and variance for up to 7 Mb of the reference genome, encompassing 108 genes (Sup Fig 3). While the *EPSPS* amplification showed the strongest selective signal on the *EPSPS*-bearing chromosome, we found distinct selective patterns associated with *EPSPS* across agricultural regions. Sweepfinder2 (35, 36) estimated the strongest amplification-related sweep signal in Walpole; the top 5% of putatively

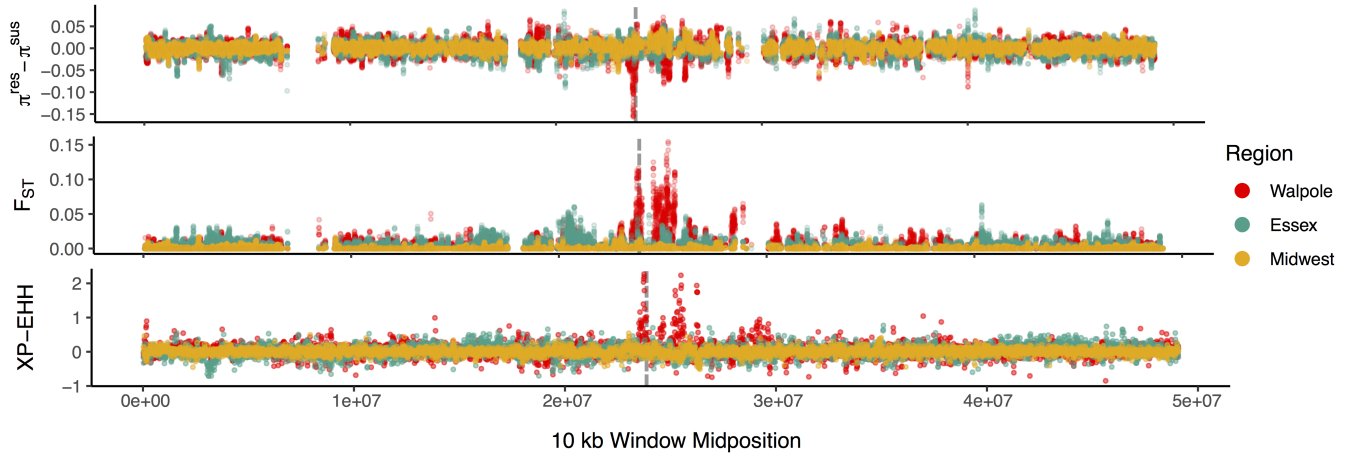


Fig. 3. Population genetic signals of selection near *EPSPS*. Comparisons of the deficit of diversity (top row), relative differentiation (middle row), and extended haplotype homozygosity (XP-EHH (34)) (bottom row) between phenotypically resistant and susceptible individuals in each agricultural region, across chromosome 5. *EPSPS* is delimited by the vertical grey dashed line, while the *EPSPS*-related amplification spans 23 - 30 Mb on chromosome 5.

selected windows experienced an estimated 50x and 100x stronger selection in phenotypically resistant individuals in Walpole compared to Essex and the Midwest, respectively (Sup Fig 4). Moreover, there was a marked reduction in genetic diversity around *EPSPS*, as well as elevated differentiation and extended haplotype homozygosity XP-EHH score (34), in resistant compared to susceptible individuals from Walpole (across the 7 Mb region: π^{res}-π^{sus} = -0.0087, F_{ST} = 0.0059, XP-EHH = 0.0472) (Fig 3). In contrast, diversity was elevated and differentiation reduced in resistant individuals from Essex and the Midwest, where the latter actually showed excess heterozygosity (Essex: π^{res}-π^{sus} = 0.0013, F_{ST} = 0.0028, XP-EHH = 0.0297; Midwest: π^{res}-π^{sus} = 0.0078, F_{ST} = 0.0006, XP-EHH = -0.0019) (Fig 3).

These differences in the estimated severity of selection and the extent of sweep signals among agricultural regions are thus likely to be driven by whether adaptation is proceeding from soft versus hard sweeps (37–40). We therefore mapped the distribution of copy number onto a maximum likelihood haplotype tree of SNPs within *EPSPS* (Fig 4), and indeed, patterns suggest that the number of origins of resistance varies considerably between agricultural regions. Whereas Walpole resistant haplotypes are highly clustered, implying a single independent origin, resistant haplotypes in Essex are scattered between susceptible haplotypes, both within and between populations. Similarly in the Midwest, *EPSPS* resistant haplotypes show soft state-level patterns, however the tree is punctuated both by clusters of resistant populations and populations where resistant individuals are spread across the tree, implying both independent evolutionary origins within and among populations in the Midwest (Fig 4).

To further assess these polymorphism-based inferences, we also looked at the similarity in the copy number profiles of the *EPSPS* amplified region; Indeed, they vary considerably across our samples (Fig 5A), suggesting multiple independent amplification events that subsequently spread through a range-wide soft selective sweep. To quantify this, we calculated for all possible pairs of resistant individuals, how well genomic coverage in the 1 Mb region surrounding *EPSPS* was correlated between them (Fig 5B). Again, the two Canadian regions showed very different patterns; coverage in individuals from Walpole island was very highly correlated (average of Spearman's ρ= 0.95), suggesting the spread of a single amplification haplotype through a hard selective sweep, while the average correlation was much lower in Essex (ρ= 0.56), even when comparing individuals from the same populations (ρ= 0.54 and 0.61), suggestive of multiple independent amplification haplotypes (Fig 4). Similar to Essex, there appeared to be multiple amplification haplotypes in the Midwest (ρ= 0.47), with evidence consistent with either hard (ρ= 0.94, 0.95, 0.93) and soft sweeps (ρ= 0.66, 0.74, 0.75) in individual populations (Fig 4).

Further investigations into these patterns of genetic differentiation and similarity in the amplification profile among agricultural regions can help to distinguish among modes of adaptation and the evolutionary mechanisms by which glyphosate resistance has spread. Although Walpole shows signs of admixture from var. *rudis*, polymorphism at *EPSPS* in Walpole is clearly differentiated from both Essex and the Midwest (Sup Fig 5), and while copy number profiles are almost perfectly correlated within Walpole, they are distinct from those found in Essex and the Midwest (Fig 5). This suggests that the evolution of glyphosate resistance in Walpole occurred independently, likely from selection on a *de novo* mutation. However, adaptive introgression of the *EPSPS* amplification into Walpole from an unsampled population is also possible. In contrast to Walpole, Essex shows low differentiation chromosome-wide and at *EPSPS* with the Midwest (Sup Fig 5), low within region copy profile correlations, but interestingly, has sporadic high correlations with a number of individuals that span many Midwestern populations (Fig 5). Distinct from these patterns, the Midwest shows high within-population correlations, where amplified individuals in these populations typically have one to a few high frequency amplification haplotypes segregating. Given the strength of within population correlations in the Midwest, and that from our demographic inference Essex appears to be

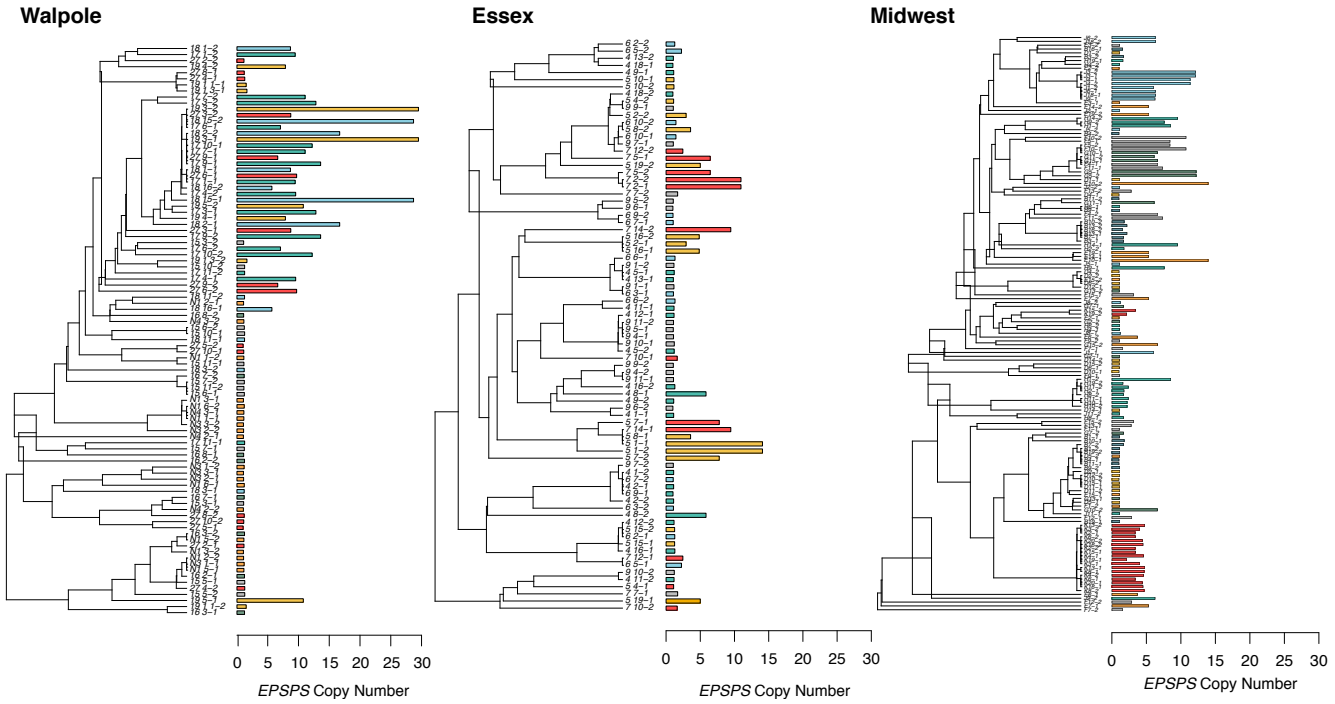


Fig. 4. Diversity of origins of *EPSPS* amplification in different agricultural regions. Haplotype tree of SNPs within the *EPSPS* gene, alongside *EPSPS* copy number by geographic region. Bars are coloured by population identity and phased haplotypes from the same individual are arbitrarily numbered -1 and -2.

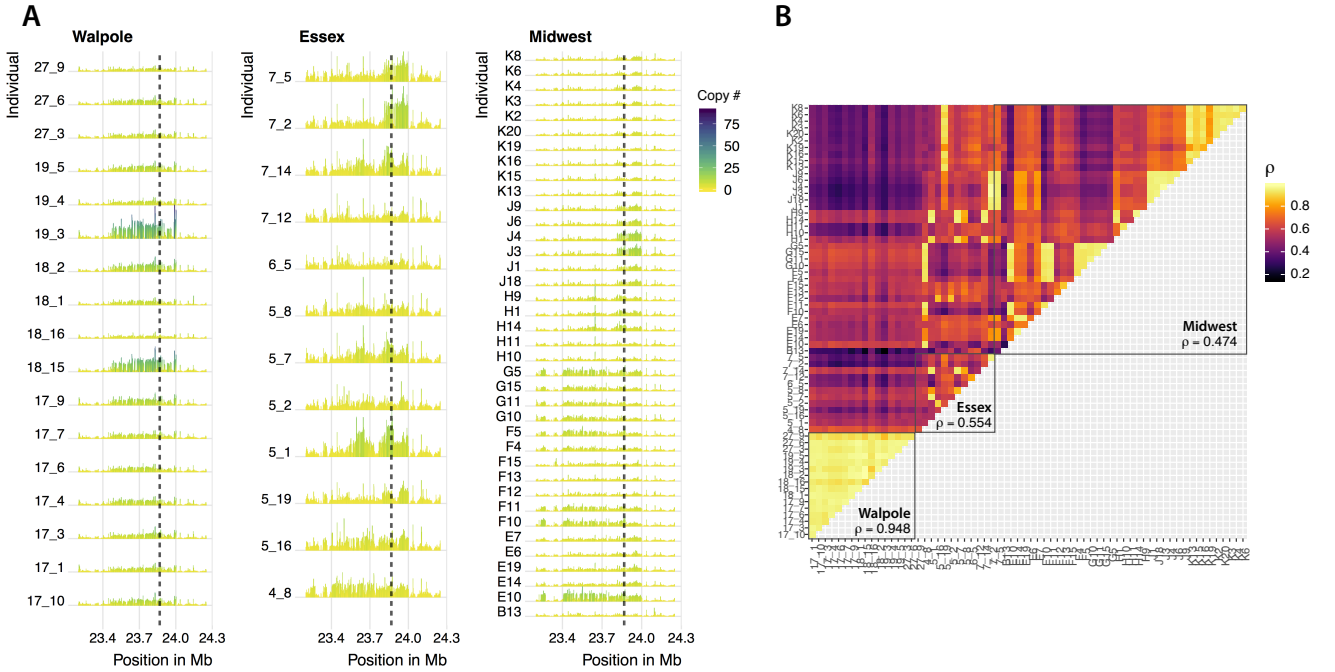


Fig. 5. A) The copy number haplotype from genomic coverage for 1 Mb around *EPSPS* (locus delimited by dashed lines) for individuals identified to carry the amplification. B) The matrix of all pairwise correlation coefficients for these same individuals and the same 1 Mb region, with each agricultural region outlined by a square and the mean coefficient across all individuals within that region provided within.

a recent seed-mediated dispersal event, the shared origins of this amplification between Essex and the Midwest is likely to have occurred via gene flow. Thus, together with results from population structure and demographic history, resistance evolution on the more agriculturally-naive var. *tuberculatus* background seems to be occurring in a mutation-limited framework, relying on evolutionary rescue via *de novo* mutation. In contrast, and as suggested in Ralph & Coop 2010 and Kreiner *et al*., 2018 (41, 42), a longer history of temporally and geographically fluctuating selection for glyphosate resistance on the var. *rudis* background in the Midwest seems to be maintaining multiple independent amplification haplotypes both within and among populations, some of which appear to have spread to Essex via gene flow.

In summary, this work highlights multiple modes of convergent evolution in the spread of glyphosate resistance, through several independent origins from new mutations, selection from recently arisen pre-existing variation, and gene flow via seed translocation. Moreover, we show that the propensity for adaptation from soft selective sweeps depends on the timescale of selection, with populations naive to agricultural environments being apparently limited to adaptation from new mutation. That agricultural adaptation of historically non-weedy lineages can occur on contemporary timescales calls for broader management strategies that encompass preventing seemingly benign weeds from establishing and adapting, regional seed containment, and local integrative control of herbicide-resistant weeds.

AUTHOR CONTRIBUTIONS

Conceptualization: Julia M. Kreiner, Stephen Wright, John Stinchcombe, Patrick J. Tranel & Detlef Weigel
Data curation: Julia M. Kreiner, Darci Ann Giacomini, Bridgit Waithaka, Felix Bemm, Christa Lanz, Julia Hildebrandt, & Julian Regalado
Filtering & validation: Julia M. Kreiner & Felix Bemm
Analysis: Julia M. Kreiner, Darci Ann Giacomini, Felix Bemm, & Julian Regalado
Methodology: Julia M. Kreiner, Stephen I. Wright
Supervision: Stephen I. Wright, John R. Stinchombe, Patrick J. Tranel, & Detlef Weigel
Visualization: Julia M. Kreiner
Writing – original draft: Julia M. Kreiner
Writing – review & editing: Julia M. Kreiner, Stephen I. Wright, John R. Stinchcombe, Detlef Weigel, Patrick J. Tranel, Darci Ann Giacomini, Peter H. Sikkema, Bridgit Waithaka, & Felix Bemm.

ACKNOWLEDGEMENTS

We thank Tyler Kent and Anna O'Brien for useful discussion and Rebecca Schwab, Fernando Rabanal and Talia Karasov for comments on the manuscript. This work was supported by NSERC Discovery Grants (SIW, JRS), NSERC EWR Steacie fellowship (SIW), Canada Research Chair (SIW), NSERC PGS-D (JMK), IMPRS Molecules to Organisms (BW), Max Planck Society and Ministry for Science, and Research and Art of Baden-Württemberg in the Regio-Research-Alliance "Yield stability in dynamic environments" (DW).

References

1. Ian Heap. Herbicide resistant weeds. In *Integrated Pest Management*, pages 281–301. Springer, Dordrecht, 2014.
2. M G Schryver, N Soltani, D C Hooker, D E Robinson, P J Tranel, and P H Sikkema. Glyphosate-resistant waterhemp (*Amaranthus tuberculatus* var. *rudis*) in ontario, canada. *Can. J. Plant Sci.*, 97(6):1057–1067, April 2017.
3. Mihai Costea, Susan E Weaver, and François 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(2):507–522, April 2005.
4. Lawrence E Steckel and Christy L Sprague. Common waterhemp (*Amaranthus rudis*) interference in corn. *Weed Sci.*, 52(3):359–364, May 2004.
5. Malcolm D Devine and Amit Shukla. Altered target sites as a mechanism of herbicide resistance. *Crop Prot.*, 19(8):881–889, September 2000.
6. Joshua S Yuan, Patrick J Tranel, and C Neal Stewart, Jr. Non-target-site herbicide resistance: a family business. *Trends Plant Sci.*, 12(1):6–13, January 2007.
7. Christophe Délye, Marie Jasieniuk, and Valérie Le Corre. Deciphering the evolution of herbicide resistance in weeds. *Trends Genet.*, 29(11):649–658, November 2013.
8. Jiaqi Guo, Chance W Riggins, Nicholas E Hausman, Aaron G Hager, Dean E Riechers, Adam S Davis, and Patrick J Tranel. Nontarget-Site resistance to ALS inhibitors in waterhemp (*amaranthus tuberculatus*). *Weed Sci.*, 63(2):399–407, 2015.
9. Vijay K Nandula, Jeffery D Ray, Daniela N Ribeiro, Z Pan, and Krishna 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(3):374–383, July 2013.
10. Laura A Chatham, Chenxi Wu, Chance W Riggins, Aaron G Hager, Bryan G Young, Gordon K Roskamp, and Patrick J Tranel. *EPSPS* gene amplification is present in the majority of glyphosate-resistant illinois waterhemp (*Amaranthus tuberculatus*) populations. *Weed Technol.*, 29(1):48–55, March 2015.
11. Laura A Chatham, Kevin W Bradley, Greg R Kruger, James R Martin, Micheal D K Owen, Dallas E Peterson, Jugulam Mithila, and Patrick J Tranel. A multistate study of the association between glyphosate resistance and *epsps* gene amplification in waterhemp (*Amaranthus tuberculatus*). *Weed Sci.*, 63(3):569–577, July 2015.
12. Lothar Lorentz, Todd A Gaines, Scott J Nissen, Philip Westra, Harry J Strek, Heinz W Dehne, Juan Pedro Ruiz-Santaella, and Roland Beffa. Characterization of glyphosate resistance in *Amaranthus tuberculatus* populations. *J. Agric. Food Chem.*, 62(32):8134–8142, August 2014.
13. Dal-Hoe Koo, Mithila Jugulam, Karthik Putta, Ivan B Cuvaca, Dallas E Peterson, Randall S Currie, Bernd Friebe, and Bikram S Gill. Gene duplication and aneuploidy trigger rapid evolution of herbicide resistance in common waterhemp. *Plant Physiol.*, 176(3):1932–1938, March 2018.
14. Dal-Hoe Koo, William T Molin, Christopher A Saski, Jiming Jiang, Karthik Putta, Mithila Jugulam, Bernd Friebe, and Bikram S Gill. Extrachromosomal circular DNA-based amplification and transmission of herbicide resistance in crop weed *Amaranthus palmeri*. *Proc. Natl. Acad. Sci. U. S. A.*, 115(13):3332–3337, March 2018.
15. Todd A Gaines, Wenli Zhang, Dafu Wang, Bekir Bukun, Stephen T Chisholm, Dale L Shaner, Scott J Nissen, William L Patzoldt, Patrick J Tranel, A Stanley Culpepper, Timothy L Grey, Theodore M Webster, William K Vencill, R Douglas Sammons, Jiming Jiang, Christopher Preston, Jan E Leach, and Philip Westra. Gene amplification confers glyphosate resistance in *Amaranthus palmeri*. *Proc. Natl. Acad. Sci. U. S. A.*, 107(3):1029–1034, January 2010.
16. Andrew Dillon, Vijay K Varanasi, Tatiana V Danilova, Dal-Hoe Koo, Sridevi Nakka, Dallas E Peterson, Patrick J Tranel, Bernd Friebe, Bikram S Gill, and Mithila Jugulam. Physical mapping of amplified copies of the 5-enolpyruvylshikimate-3-phosphate synthase gene in glyphosate-resistant *Amaranthus tuberculatus*. *Plant Physiol.*, 173(2):1226–1234, February 2017.
17. Eric L Patterson, Dean J Pettinga, Karl Ravet, Paul Neve, and Todd A Gaines. Glyphosate resistance and *epsps* gene duplication: convergent evolution in multiple plant species. *J. Hered.*, 109(2):117–125, February 2018.
18. V Fugère, M P Hébert, N B Costa, C C Y Xu, R D H Barrett, and others. Community rescue in experimental phytoplankton communities facing severe herbicide pollution. *bioRxiv*, 2018.
19. Adam Kuester, Shu-Mei Chang, and Regina S Baucom. The geographic mosaic of herbicide resistance evolution in the common morning glory, *Ipomoea purpurea*: Evidence for resistance hotspots and low genetic differentiation across the landscape. *Evol. Appl.*, 8(8): 821–833, September 2015.
20. Anita Küpper, Harish K Manmathan, Darci Giacomini, Eric L Patterson, William B McCloskey, and Todd A Gaines. Population genetic structure in glyphosate-resistant and -susceptible palmer amaranth (*Amaranthus palmeri*) populations using genotyping-by-sequencing (GBS). *Front. Plant Sci.*, 9:29, January 2018.
21. Chenxi Wu, Adam S Davis, and Patrick J Tranel. Limited fitness costs of herbicide-resistance traits in *Amaranthus tuberculatus* facilitate resistance evolution. *Pest Manag. Sci.*, 74(2):293–301, February 2018.
22. A J Price, K S Balkcom, S A Culpepper, J A Kelton, R L Nichols, and H Schomberg. Glyphosate-resistant palmer amaranth: A threat to conservation tillage. *J. Soil Water Conserv.*, 66(4):265–275, July 2011.
23. Markus G Stetter and Karl 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, April 2017.
24. Felipe A Simão, Robert M Waterhouse, Panagiotis Ioannidis, Evgenia V Kriventseva, and Evgeny M Zdobnov. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19):3210–3212, October 2015.
25. D J Lightfoot, D E Jarvis, T Ramaraj, R Lee, E N Jellen, and P J Maughan. Single-molecule sequencing and Hi-C-based proximity-guided assembly of amaranth (*Amaranthus hypochondriacus*) chromosomes provide insights into genome evolution. *BMC Biol.*, 15(1): 74, August 2017.
26. Jun Chen, Sylvain Glémin, and Martin Lascoux. Genetic diversity and the efficacy of purifying selection across plant and animal species. *Mol. Biol. Evol.*, 34(6):1417–1428, June 2017.
27. Katherine E Waselkov and Kenneth M Olsen. Population genetics and origin of the native north american agricultural weed waterhemp (*Amaranthus tuberculatus*; amaranthaceae). *Am. J. Bot.*, 101(10):1726–1736, October 2014.
28. Jonathan Sauer. Recent migration and evolution of the dioecious amaranths. *Evolution*, 11 (1):11–31, March 1957.
29. Simon H Martin, John W Davey, and Chris D Jiggins. Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Mol. Biol. Evol.*, 32(1):244–257, January 2015.
30. Jaime A Farmer, Elisabeth B Webb, Robert A Pierce, 2nd, and Kevin W Bradley. Evaluating the potential for weed seed dispersal based on waterfowl consumption and seed viability. *Pest Manag. Sci.*, 73(12):2592–2603, December 2017.
31. Katherine Waselkov. *Population genetics and phylogenetic context of weed evolution in the genus Amaranthus: Amaranthaceae)*. PhD thesis, Washington University in St. Louis, 2013.
32. Chun-Hai Dong, Bethany K Zolman, Bonnie Bartel, Byeong-Ha Lee, Becky Stevenson, Manu Agarwal, and Jian-Kang Zhu. Disruption of *Arabidopsis* CHY1 reveals an important role of metabolic status in plant cold stress signaling. *Mol. Plant*, 2(1):59–72, January

2009.

33. Peter R Lange, Peter J Eastmond, Kathryn Madagan, and Ian A Graham. An *Arabidopsis* mutant disrupted in valine catabolism is also compromised in peroxisomal fatty acid β-oxidation. *FEBS Lett.*, 571(1-3):147–153, 2004.
34. Pardis C Sabeti, Patrick Varilly, Ben Fry, Jason Lohmueller, Elizabeth Hostetter, Chris Cotsapas, Xiaohui Xie, Elizabeth H Byrne, Steven A McCarroll, Rachelle Gaudet, Stephen F Schaffner, Eric S Lander, International HapMap Consortium, Kelly A Frazer, Dennis G Ballinger, David R Cox, David A Hinds, Laura L Stuve, Richard A Gibbs, John W Belmont, Andrew Boudreau, Paul Hardenbol, Suzanne M Leal, Shiran Pasternak, David A Wheeler, Thomas D Willis, Fuli Yu, Huanming Yang, Changqing Zeng, Yang Gao, Haoran Hu, Weitao Hu, Chaohua Li, Wei Lin, Siqi Liu, Hao Pan, Xiaoli Tang, Jian Wang, Wei Wang, Jun Yu, Bo Zhang, Qingrun Zhang, Hongbin Zhao, Hui Zhao, Jun Zhou, Stacey B Gabriel, Rachel Barry, Brendan Blumenstiel, Amy Camargo, Matthew Defelice, Maura Faggart, Mary Goyette, Supriya Gupta, Jamie Moore, Huy Nguyen, Robert C Onofrio, Melissa Parkin, Jessica Roy, Erich Stahl, Ellen Winchester, Liuda Ziaugra, David Altshuler, Yan Shen, Zhijian Yao, Wei Huang, Xun Chu, Yungang He, Li Jin, Yangfan Liu, Yayun Shen, Weiwei Sun, Haifeng Wang, Yi Wang, Ying Wang, Xiaoyan Xiong, Liang Xu, Mary M Y Waye, Stephen K W Tsui, Hong Xue, J Tze-Fei Wong, Luana M Galver, Jian-Bing Fan, Kevin Gunderson, Sarah S Murray, Arnold R Oliphant, Mark S Chee, Alexandre Montpetit, Fanny Chagnon, Vincent Ferretti, Martin Leboeuf, Jean-François Olivier, Michael S Phillips, Stéphanie Roumy, Clémentine Sallée, Andrei Verner, Thomas J Hudson, Pui-Yan Kwok, Dongmei Cai, Daniel C Koboldt, Raymond D Miller, Ludmila Pawlikowska, Patricia Taillon-Miller, Ming Xiao, Lap-Chee Tsui, William Mak, You Qiang Song, Paul K H Tam, Yusuke Nakamura, Takahisa Kawaguchi, Takuya Kitamoto, Takashi Morizono, Atsushi Nagashima, Yozo Ohnishi, Akihiro Sekine, Toshihiro Tanaka, Tatsuhiko Tsunoda, Panos Deloukas, Christine P Bird, Marcos Delgado, Emmanouil T Dermitzakis, Rhian Gwilliam, Sarah Hunt, Jonathan Morrison, Don Powell, Barbara E Stranger, Pamela Whittaker, David R Bentley, Mark J Daly, Paul I W de Bakker, Jeff Barrett, Yves R Chretien, Julian Maller, Steve McCarroll, Nick Patterson, Itsik Pe'er, Alkes Price, Shaun Purcell, Daniel J Richter, Pardis Sabeti, Richa Saxena, Stephen F Schaffner, Pak C Sham, Patrick Varilly, David Altshuler, Lincoln D Stein, Lalitha Krishnan, Albert Vernon Smith, Marcela K Tello-Ruiz, Gudmundur A Thorisson, Aravinda Chakravarti, Peter E Chen, David J Cutler, Carl S Kashuk, Shin Lin, Gonçalo R Abecasis, Weihua Guan, Yun Li, Heather M Munro, Zhaohui Steve Qin, Daryl J Thomas, Gilean McVean, Adam Auton, Leonardo Bottolo, Niall Cardin, Susana Eyheramendy, Colin Freeman, Jonathan Marchini, Simon Myers, Chris Spencer, Matthew Stephens, Peter Donnelly, Lon R Cardon, Geraldine Clarke, David M Evans, Andrew P Morris, Bruce S Weir, Tatsuhiko Tsunoda, Todd A Johnson, James C Mullikin, Stephen T Sherry, Michael Feolo, Andrew Skol, Houcan Zhang, Changqing Zeng, Hui Zhao, Ichiro Matsuda, Yoshimitsu Fukushima, Darryl R Macer, Eiko Suda, Charles N Rotimi, Clement A Adebamowo, Ike Ajayi, Toyin Aniagwu, Patricia A Marshall, Chibuzor Nkwodimmah, Charmaine D M Royal, Mark F Leppert, Missy Dixon, Andy Peiffer, Renzong Qiu, Alastair Kent, Kazuto Kato, Norio Niikawa, Isaac F Adewole, Bartha M Knoppers, Morris W Foster, Ellen Wright Clayton, Jessica Watkin, Richard A Gibbs, John W Belmont, Donna Muzny, Lynne Nazareth, Erica Sodergren, George M Weinstock, David A Wheeler, Imtaz Yakub, Stacey B Gabriel, Robert C Onofrio, Daniel J Richter, Liuda Ziaugra, Bruce W Birren, Mark J Daly, David Altshuler, Richard K Wilson, Lucinda L Fulton, Jane Rogers, John Burton, Nigel P Carter, Christopher M Clee, Mark Griffiths, Matthew C Jones, Kirsten McLay, Robert W Plumb, Mark T Ross, Sarah K Sims, David L Willey, Zhu Chen, Hua Han, Le Kang, Martin Godbout, John C Wallenburg, Paul L'Archevêque, Guy Bellemare, Koji Saeki, Hongguang Wang, Daochang An, Hongbo Fu, Qing Li, Zhen Wang, Renwu Wang, Arthur L Holden, Lisa D Brooks, Jean E McEwen, Mark S Guyer, Vivian Ota Wang, Jane L Peterson, Michael Shi, Jack Spiegel, Lawrence M Sung, Lynn F Zacharia, Francis S Collins, Karen Kennedy, Ruth Jamieson, and John Stewart. Genome-wide detection and characterization of positive selection in human populations. *Nature*, 449(7164):913–918, October 2007.
35. Michael DeGiorgio, Christian D Huber, Melissa J Hubisz, Ines Hellmann, and Rasmus Nielsen. SweepFinder2: increased sensitivity, robustness and flexibility. *Bioinformatics*, 32(12):1895–1897, June 2016.
36. Christian D Huber, Michael DeGiorgio, Ines Hellmann, and Rasmus Nielsen. Detecting recent selective sweeps while controlling for mutation rate and background selection. *Mol. Ecol.*, 25(1):142–156, January 2016.
37. Joachim Hermisson and Pleuni S Pennings. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, 169(4):2335–2352, April 2005.
38. Pleuni S Pennings and Joachim Hermisson. Soft sweeps III: the signature of positive selection from recurrent mutation. *PLoS Genet.*, 2(12):e186, December 2006.
39. Pleuni S Pennings and Joachim Hermisson. Soft sweeps II—Molecular population genetics of adaptation from recurrent mutation or migration. *Mol. Biol. Evol.*, 23(5):1076–1084, May 2006.
40. Joachim Hermisson and Pleuni S Pennings. Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods Ecol. Evol.*, 8(6):700–716, June 2017.
41. Julia M Kreiner, John R Stinchcombe, and Stephen I Wright. Population genomics of herbicide resistance: adaptation via evolutionary rescue. *Annu. Rev. Plant Biol.*, 69:611–635, April 2018.
42. Peter Ralph and Graham Coop. Parallel adaptation: one or many waves of advance of an advantageous allele? *Genetics*, 186(2):647–668, October 2010.

Methods

Plant Collections. Seed was collected from Ontario natural populations and agricultural fields in the fall of 2016, and Mid-western populations in 2010 initially for investigation in *Chatham et al.*, 2015 (1). Agricultural fields that exhibited poor control of *A. tuberculatus* were selected for sampling and thus are biased towards particularly high levels of glyphosate resistance, and do not accurately represent levels of resistance across randomly sampled populations across the range.

High Molecular Weight DNA Extraction. High molecular weight (HMW) DNA was extracted from the leaf tissue of a single 28-day-old glyphosate-resistant male *A. tuberculatus* plant from the Midwest United States using a modified version of the Doyle and Doyle nuclei isolation protocol (2).

Nuclei isolation was carried out by incubating 30 g of ground leaf tissue in a buffer comprising tris(hydroxymethyl)aminomethane, potassium chloride, ethylenediaminetetraacetic acid, sucrose, spermidine and spermine tetrahydrochloride (Sigma-Aldrich, MO, USA). The homogenate was subsequently filtered using miracloth and precipitated by centrifugation. G2 lysis buffer, RNase A and Proteinase K (Qiagen, Venlo, Netherlands) were then added prior to an overnight incubation at 50°C and centrifugation at 4°C. The supernatant containing the DNA solution was then added to an equilibrated Qiagen genomic tip 100 (Qiagen, Venlo, Netherlands). Following this, clean genomic DNA was eluted and precipitated using isopropanol. Finally, high molecular weight DNA was isolated by DNA spooling.

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

Lucigen PCR-free Library Preparation and Sequencing. DNA from natural and agricultural *A. tuberculatus* populations sampled from the Midwest United States and Ontario was fragmented to a 350 bp insert size using a Covaris S2 Focused Ultrasonicator (Covaris, MA, USA). Subsequent end-repair, A-tailing, Lucigen adaptor ligation and size-selection was performed using the Lucigen NxSeq® AMPFree Low DNA Library Kit (Lucigen, WI, USA). Libraries were quantified using the Qubit 2.0 (Life Technologies, CA, USA) while library profiles were analysed using a Bioanalyzer High Sensitivity Chip on an Agilent Bioanalyzer 2100 (Agilent Technologies, CA, USA). The libraries were then sequenced to a coverage depth of 10X on an Illumina HiSeq 3000 instrument using a HiSeq 3000/4000 SBS kit and paired-end 150 base read chemistry.

Genome assembly & haplotype merging. The genome was assembled from 58 Gb of long read data using Canu (version 1.6; genomeSize=544m; other parameters default) (3). 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) (4). Polished contigs were repeat masked using WindowMasker (version 1.0.0; -checkdup; other parameters default) (5). Repeat-masked contigs were screened for misjoints and subjected to haplotype merging using HaploMerger2 (commit 95f8589; identity=80, other parameters default (6). 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) (7) taking only the best contig-to-contig alignments into account. The final assembly was finished against the chromosome-resolved *A. hypochondriacus* genome (8) using reveal finish (commit 98d3ad1; –fixedgapsize –gapsize 15,000; other parameters default) (9). The 16 resulting pseudo chromosomes represented 99.6% of our original assembly and were used for all chromosome-wide scans, such as sweep signal detection

Assembly, SNP calling, and gene annotation. We used freebayes (10) parallel to call SNPs jointly on all samples. For whole genome analyses, we used a thoroughly filtered SNP set following the guidelines of Fang 2014 and dDocent (11, 12) adapted for whole genome data: sites were removed based on missing data (>80%), complexity, indels, allelic bias (<0.25 & >0.75), whether there was a discrepancy in paired status of reads supporting reference or alternate alleles, mapping quality (QUAL < 30, representing sites with greater than a 1/1000 error rate), and lastly, individuals with excess missing data (>5%) were dropped. This led to a final, high confident set of 10,280,132 SNPs. For *EPSPS* specific analyses and genome wide investigations that required invariant sites, we recalled SNPs with samtools(V1.7) and bwa-mem (V0.7.17). Bams were sorted and duplicates marked with sambamba (V0.6.6), while read groups were added with picard (V2.17.11). Sites were minimally

filtered on mapping quality and missing data, (keeping only sites with MQ > 30 & < 20% missing data), so that we did not bias our diversity estimates by preferentially retaining invariant or variant sites.

We performed gene annotation on both our final assembly and the hypochondriacus-finished pseudoassembly using the MAKER annotation pipeline. *A. tuberculatus*-specific repeats were identified using RepeatModeler (v1.0.11; (13)), combined with the RepBase repeat library, and masked with RepeatMasker (v4.0.7; (14)). This repeat-masked genome was then run through MAKER (v2.31.8; (15)), using EST evidence from an *A. tuberculatus* transcriptome assembly (16) and protein homology evidence from *A. hypochondriacus* (17). The gene models were further annotated using InterProScan (v69.0; (18)), resulting in a total of 30,771 genes and 40,766 transcripts with a mean transcript length of 1245 bp. The mean annotation edit distance (AED) score was 0.21 and 98.1% of the gene predictions had an AED score of less than 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 5-7 leaf stage with 1,260 g active ingredient per hectare glyphosate (WeatherMax 4.5 L, Monsanto, Chesterfield, MO). Fourteen days after treatment, plants were rated visually on a scale of 0 (highly sensitive) to 5 (no injury). Plants rated with a 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 gDNA extraction. Tissue from plants rated as highly glyphosate-resistant or susceptible were selected from each population for genomic DNA extraction using a modified CTAB method (2).

Copy number estimates. The scaled coverage haplotype 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 and regions of low coverage (<3X), which should represent the coverage of single-copy genes.

Structure, demographic modelling & summary statistics. In order 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 there to be 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 later was used as input for demographic modelling.

Our two-population demographic model of *A. tuberculatus* modelled the split between the *A. tuberculatus* var. *tuberculatus* and var. *rudis* subspecies, by collapsing individuals into one of the two lineages by their predominant ancestry as identified in our STRUCTURE analyses. This was estimated in ai (V1.7.0)(19) using the pipeline available on https://github.com/dportik/dadi_pipeline (20). 1D and 2D site frequency spectrums were estimated using the program easySFS (https://github.com/isaacovercast/easySFS), where samples where our SFS were projected downwards to exclude missing data, while maximizing the total number of sites and individuals. We ensured that the log-likelihood of our parameter set had optimized by iterating the analysis over four rounds of increasing reps, from 10 to 40. We tested a set of 20 diversification models, that varied 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 with no gene flow, followed by size change with asymmetrical gene flow. estimated 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*Na generations) (*T1*), the scaled time between the secondary contact and present (*T2*), size of population 1 after time interval (*nu1b*), size of population 2 after time interval (*nu2b*), migration from pop 2 to pop 1 (*2*Na*m12*), and migration from pop 1 to pop 2 (*m21*). N_e was calculated by subbing the per site theta estimate (after controlling for the effective sequence length to account for losses in the alignment and missed or filtered calls), and the *A. thaliana* mutation rate ($7*10^{-9}$) (21) into the equation $\theta = 4N_e\mu$.

We used *PLINK* (V1.9) to perform a PCA of genotypes from our final freebayes SNP set after thinning for linkage disequilibrium, *STRUCTURE* (V2.3.4) (22) to estimate admixture across populations, and treemix (V3) (23) 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 10 kb windows, overlapping by 1 kb. To estimate the proportion of introgression of var. *rudis* ancestry into Walpole agricultural populations (*f*), we also used the genomics general pipeline, however using a window sizes of 100 kb with 10 kb overlaps, to minimize stochasticity in these estimates due to a low number of SNPs. Specifically, we looked at the proportion of introgression from Essex (P3) into Walpole (P2), relative to natural populations (P1), as well as the proportion of introgression from Midwestern populations (P3) into Essex (P2), relative to natural populations (P1), using *A. hypochondriacus* as an outgroup. We then used a blocked jacknife to attain confidence interval estimates for *f*, using a block size of 1 Mb. To use A. hypochondriacus as the outgroup, we aligned the hypochondriacus genome to our *A. tuberculatus* pseudoreference with *LASTZ* (24). For

the outlier analysis of putative genes underlying contemporary agricultural adaptation in Walpole, we performed a regression of within Walpole diversity against between Walpole-Natural population differentiation. We then classified windows as outliers that had the top 1% of extreme values of differentiation for a given level of diversity, which should represent regions in the genome, specific to Walpole, that have recently undergone positive selection. A GO enrichment test was then performed for these outlier regions, after finding their intersect annotated *A. tuberculatus* genes, and their orthologues in *A. thaliana* using orthofinder (25).

Detecting selective sweeps & estimating recombination rate. To detect differences in the strength and breadth of sweep signal associated with selection from glyphosate herbicides across geographic regions, we used SNPs called from the psuedoassembly of our *A. tuberculatus* reference mapped onto the fully resolved *A. hypochondriacus* genome (as described above, with same calling procedures as in the SNP calling section). 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* (26) to estimate variable recombination rate independently across all 16 chromosomes of the pseudoassembly, using a precomputed lookup table for a theta 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`.

We ran *BEAGLE* (V4.0) (27) to phased haplotypes on chromosome 5, where the *EPSPS* gene is localized. These phased haplotypes were used for the haplotype-homozygosity based sweep analyses, XP-EHH (28), calculated based on the difference in haplotype homozygosity between resistant and susceptible individuals for each geographic region after controlling for recombination rate, all of which was implemented in selscan (29). Phased haplotypes were also used to calculate a maximum likelihood tree for the 235 SNPs that fell within the *EPSPS* gene. For each tree, we realigned sequences before bootstrapping 1,000 replicates of our haplotree with clustal omega (30). In contrast to haplotype-based methods that required phased data, we also ran Sweepfinder2 (31, 32) 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. Lastly, 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*, for the length of the most proximal, continuous segment of the amplification.

References

1. Laura A Chatham, Chenxi Wu, Chance W Riggins, Aaron G Hager, Bryan G Young, Gordon K Roskamp, and Patrick J Tranel. *EPSPS* gene amplification is present in the majority of glyphosate-resistant illinois waterhemp (*Amaranthus tuberculatus*) populations. *Weed Technol.*, 29(1):48–55, March 2015.
2. J Doyle and J L Doyle. Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochem Bull*, 19(11):11–15, 1987.
3. Sergey Koren, Brian P Walenz, Konstantin Berlin, Jason R Miller, Nicholas H Bergman, and Adam M Phillippy. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.*, 27(5):722–736, May 2017.
4. Bruce J Walker, Thomas Abeel, Terrance Shea, Margaret Priest, Amr Abouelliel, Sharadha Sakthikumar, Christina A Cuomo, Qiandong Zeng, Jennifer Wortman, Sarah K Young, and Ashlee M Earl. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, 9(11):e112963, November 2014.
5. Aleksandr Morgulis, E Michael Gertz, Alejandro A Schäffer, and Richa Agarwala. WindowMasker: window-based masker for sequenced genomes. *Bioinformatics*, 22(2):134–141, January 2006.
6. Shengfeng Huang, Mingjing Kang, and Anlong Xu. HaploMerger2: rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. *Bioinformatics*, 33(16):2577–2579, August 2017.
7. Heng Li. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, May 2018.
8. D J Lightfoot, D E Jarvis, T Ramaraj, R Lee, E N Jellen, and P J Maughan. Single-molecule sequencing and Hi-C-based proximity-guided assembly of amaranth (*Amaranthus hypochondriacus*) chromosomes provide insights into genome evolution. *BMC Biol.*, 15(1):74, August 2017.
9. Jasper Linthorst, Marc Hulsman, Henne Holstege, and Marcel Reinders. Scalable multi whole-genome alignment using recursive exact matching. July 2015.
10. Erik Garrison and Gabor Marth. Haplotype-based variant detection from short-read sequencing. July 2012.
11. Han Fang. Towards better understanding of artifacts in variant calling from high-Coverage samples. 2014.
12. Jonathan B Puritz, Christopher M Hollenbeck, and John R Gold. ddocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*, 2:e431, June 2014.
13. A F A Smit and R Hubley. RepeatModeler open-1.0. *Available fom http://www. repeatmasker. org*, 2008.
14. Smit, AFA, Hubley, R & Green, P. RepeatMasker open-4.0, 2013.
15. Brandi L Cantarel, Ian Korf, Sofia M C Robb, Genis Parra, Eric Ross, Barry Moore, Carson Holt, Alejandro Sánchez Alvarado, and Mark Yandell. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.*, 18(1):188–196, January 2008.
16. Chance W Riggins, Yanhui Peng, C Neal Stewart, Jr, and Patrick 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(10):1042–1052, October 2010.
17. J. W. Clouse, D. Adhikary, J. T. Page, T. Ramaraj, M. K. Deyholos, J. A. Udall, J. A. Udall, D. J. Fairbanks, E. N. Jellen, and P. J. Maughan. The amaranth genome: genome, transcriptome, and physical map assembly.
18. Philip Jones, David Binns, Hsin-Yu Chang, Matthew Fraser, Weizhong Li, Craig McAnulla, Hamish McWilliam, John Maslen, Alex Mitchell, Gift Nuka, Sebastien Pesseat, Antony F Quinn, Amaia Sangrador-Vegas, Maxim Scheremetjew, Siew-Yit Yong, Rodrigo Lopez, and Sarah Hunter. InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30(9):1236–1240, May 2014.
19. Ryan N Gutenkunst, Ryan D Hernandez, Scott H Williamson, and Carlos D Bustamante. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.*, 5(10):e1000695, October 2009.
20. Daniel M Portik, Adam D Leaché, Danielle Rivera, Michael F Barej, Marius Burger, Mareike Hirschfeld, Mark-Oliver Rödel, David C Blackburn, and Matthew K Fujita. Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Mol. Ecol.*, 26(19):5245–5263, 2017.
21. S Ossowski, K Schneeberger, J I Lucas-Lledo, N Warthmann, R M Clark, R G Shaw, D Weigel, and M Lynch. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science*, 327(5961):92–94, 2010.
22. J K Pritchard, M Stephens, and P Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959, June 2000.
23. Joseph K Pickrell and Jonathan K Pritchard. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.*, 8(11):e1002967, November 2012.
24. Robert S Harris. Improved pairwise alignmnet of genomic DNA. 2007.
25. David M Emms and Steven Kelly. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.*, 16:157, August 2015.
26. A Auton and G McVean. Recombination rate estimation in the presence of hotspots. *Genome Res.*, 17(8):1219–1227, 2007.
27. Sharon R Browning and Brian L Browning. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.*, 81(5):1084–1097, November 2007.
28. Pardis C Sabeti, Patrick Varilly, Ben Fry, Jason Lohmueller, Elizabeth Hostetter, Chris Cotsapas, Xiaohui Xie, Elizabeth H Byrne, Steven A McCarroll, Rachelle Gaudet, Stephen F Schaffner, Eric S Lander, International HapMap Consortium, Kelly A Frazer, Dennis G Ballinger, David R Cox, David A Hinds, Laura L Stuve, Richard A Gibbs, John W Belmont, Andrew Boudreau, Paul Hardenbol, Suzanne M Leal, Shiran Pasternak, David A Wheeler, Thomas D Willis, Fuli Yu, Huanming Yang, Changqing Zeng, Yang Gao, Haoran Hu, Weitao Hu, Chaohua Li, Wei Lin, Siqi Liu, Hao Pan, Xiaoli Tang, Jian Wang, Wei Wang, Jun Yu, Bo Zhang, Qingrun Zhang, Hongbin Zhao, Hui Zhao, Jun Zhou, Stacey B Gabriel, Rachel Barry, Brendan Blumenstiel, Amy Camargo, Matthew Defelice, Maura Faggart, Mary Goyette, Supriya Gupta, Jamie Moore, Huy Nguyen, Robert C Onofrio, Melissa Parkin, Jessica Roy, Erich Stahl, Ellen Winchester, Liuda Ziaugra, David Altshuler, Yan Shen, Zhijian Yao, Wei Huang, Xun Chu, Yungang He, Li Jin, Yangfan Liu, Yayun Shen, Weiwei Sun, Haifeng Wang, Yi Wang, Ying Wang, Xiaoyan Xiong, Liang Xu, Mary M Y Waye, Stephen K W Tsui, Hong Xue, J Tze-Fei Wong, Luana M Galver, Jian-Bing Fan, Kevin Gunderson, Sarah S Murray, Arnold R Oliphant, Mark S Chee, Alexandre Montpetit, Fanny Chagnon, Vincent Ferretti, Martin Leboeuf, Jean-François Olivier, Michael S Phillips, Stéphanie Roumy, Clémentine Sallée, Andrei Verner, Thomas J Hudson, Pui-Yan Kwok, Dongmei Cai, Daniel C Koboldt, Raymond D Miller, Ludmila Pawlikowska, Patricia Taillon-Miller, Ming Xiao, Lap-Chee Tsui, William Mak, You Qiang Song, Paul K H Tam, Yusuke Nakamura, Takahisa Kawaguchi, Takuya Kitamoto, Takashi Morizono, Atsushi Nagashima, Yozo Ohnishi, Akihiro Sekine, Toshihiro Tanaka, Tatsuhiko Tsunoda, Panos Deloukas, Christine P Bird, Marcos Delgado, Emmanouil T Dermitzakis, Rhian Gwilliam, Sarah Hunt, Jonathan Morrison, Don Powell, Barbara E Stranger, Pamela Whittaker, David R Bentley, Mark J Daly, Paul I W de Bakker, Jeff Barrett, Yves R Chretien, Julian Maller, Steve McCarroll, Nick Patterson, Itsik Pe'er, Alkes Price, Shaun Purcell, Daniel J Richter, Pardis Sabeti, Richa Saxena, Stephen F Schaffner, Pak C Sham, Patrick Varilly, David Altshuler, Lincoln D Stein, Lalitha Krishnan, Albert Vernon Smith, Marcela K Tello-Ruiz, Gudmundur A Thorisson, Aravinda Chakravarti, Peter E Chen, David J Cutler, Carl S Kashuk, Shin Lin, Gonçalo R Abecasis, Weihua Guan, Yun Li, Heather M Munro, Zhaohui Steve Qin, Daryl J Thomas, Gilean McVean, Adam Auton, Leonardo Bottolo, Niall Cardin, Susana Eyheramendy, Colin Freeman, Jonathan Marchini, Simon Myers, Chris Spencer, Matthew Stephens, Peter Donnelly, Lon R Cardon, Geraldine Clarke, David M Evans, Andrew P Morris, Bruce S Weir, Tatsuhiko Tsunoda, Todd A Johnson, James C Mullikin, Stephen T Sherry, Michael Feolo, Andrew Skol, Houcan Zhang, Changqing Zeng, Hui Zhao, Ichiro Matsuda, Yoshimitsu Fukushima, Darryl R Macer, Eiko Suda, Charles N Rotimi, Clement A Adebamowo, Ike Ajayi, Toyin Aniagwu, Patricia A Marshall, Chibuzor Nkwodimmah, Charmaine D M Royal, Mark F Leppert, Missy Dixon, Andy Peiffer, Renzong Qiu, Alastair Kent, Kazuto Kato, Norio Niikawa, Isaac F Adewole, Bartha M Knoppers, Morris W Foster, Ellen Wright Clayton, Jessica Watkin, Richard A Gibbs, John W Belmont, Donna Muzny, Lynne Nazareth, Erica Sodergren, George M Weinstock, David A Wheeler, Imtaz Yakub, Stacey B Gabriel, Robert C Onofrio, Daniel J Richter, Liuda Ziaugra, Bruce W Birren, Mark J Daly, David Altshuler, Richard K Wilson, Lucinda L Fulton, Jane Rogers, John Burton, Nigel P Carter, Christopher M Clee, Mark Griffiths, Matthew C Jones, Kirsten McLay, Robert W Plumb, Mark T Ross, Sarah K Sims, David L Willey, Zhu Chen, Hua Han, Le Kang, Martin Godbout, John C Wallenburg, Paul L'Archevêque, Guy Bellemare, Koji Saeki, Hongguang Wang, Daochang An, Hongbo Fu, Qing Li, Zhen Wang, Renwu Wang, Arthur L Holden, Lisa D Brooks, Jean E McEwen, Mark S Guyer, Vivian Ota Wang, Jane L Peterson, Michael Shi, Jack Spiegel, Lawrence M Sung, Lynn F Zacharia, Francis S Collins, Karen Kennedy, Ruth Jamieson, and John Stewart. Genome-wide detection and characterization of positive selection in human populations. *Nature*, 449(7164):913–918, October 2007.
29. Zachary A Szpiech and Ryan D Hernandez. selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Mol. Biol. Evol.*, 31(10):2824–2827, October 2014.
30. Fabian Sievers, Andreas Wilm, David Dineen, Toby J Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, Hamish McWilliam, Michael Remmert, Johannes Söding, Julie D Thompson, and Desmond G Higgins. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol. Syst. Biol.*, 7:539, October 2011.
31. Christian D Huber, Michael DeGiorgio, Ines Hellmann, and Rasmus Nielsen. Detecting recent selective sweeps while controlling for mutation rate and background selection. *Mol. Ecol.*, 25(1): 142–156, January 2016.
32. Michael DeGiorgio, Christian D Huber, Melissa J Hubisz, Ines Hellmann, and Rasmus Nielsen. SweepFinder2: increased sensitivity, robustness and flexibility. *Bioinformatics*, 32(12):1895–1897, June 2016.

Supplemental Tables & Figures

Sup Table 1. Metrics of the raw and the haplotype-reduced reference genome assemblies

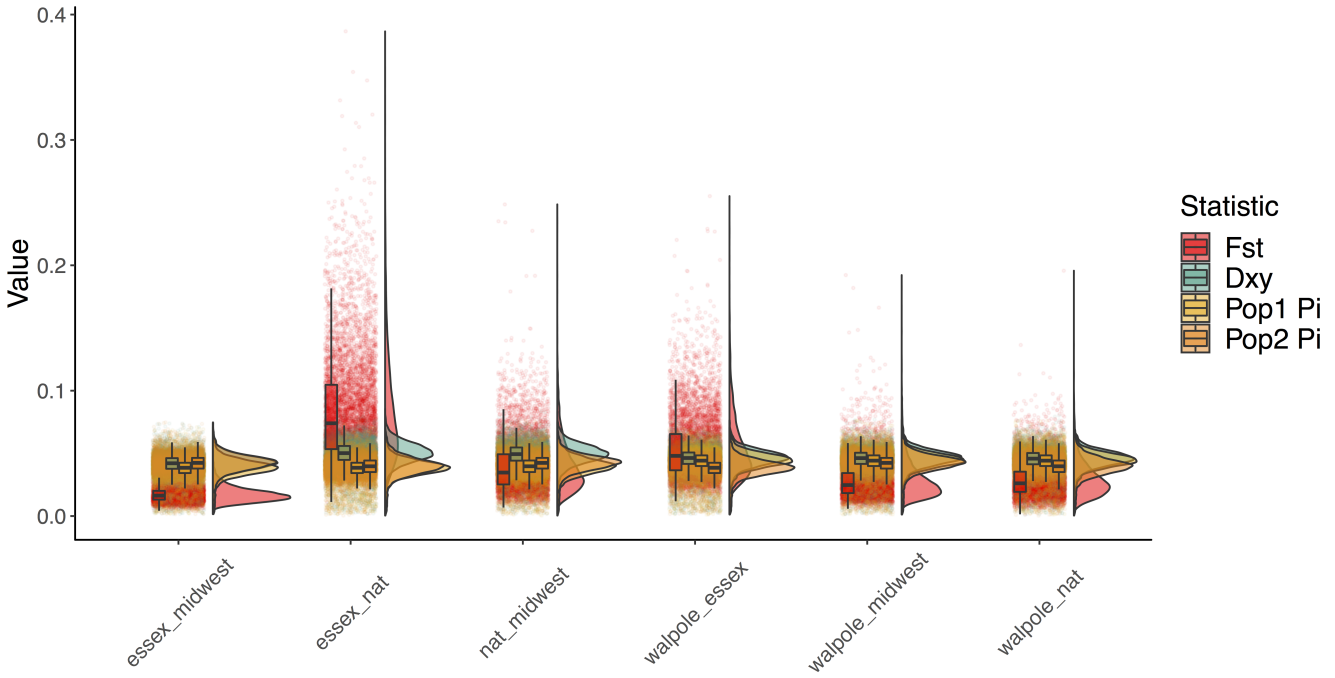
Assembly	Total Size (Mb)	Sequence (#)	N50 (bp)	Longest sequence (bp)	Completeness (%)	Duplicates (%)
Raw Assembly	1,159,758,700	4,207	905,938	9,167,955	90	71
Haploid Assembly	663,660,067	2,514	1,738,871	13,655,724	87	6

Sup Table 2. Correlation of PC1 and PC2 with both longitude and latitude. From four separately run ANOVAs.

	Longitude		
PC1	r2=0.05279	p=0.00190	F1,160=9.073
PC2	r2=0.7685	p=<2.2e-16	F1,160=535.3
	Latitude		
PC1	r2=0.08058	p=0.0001482	F1,160=15.11
PC2	r2=0.6025	p=<2.2e-16	F1,160=245.1

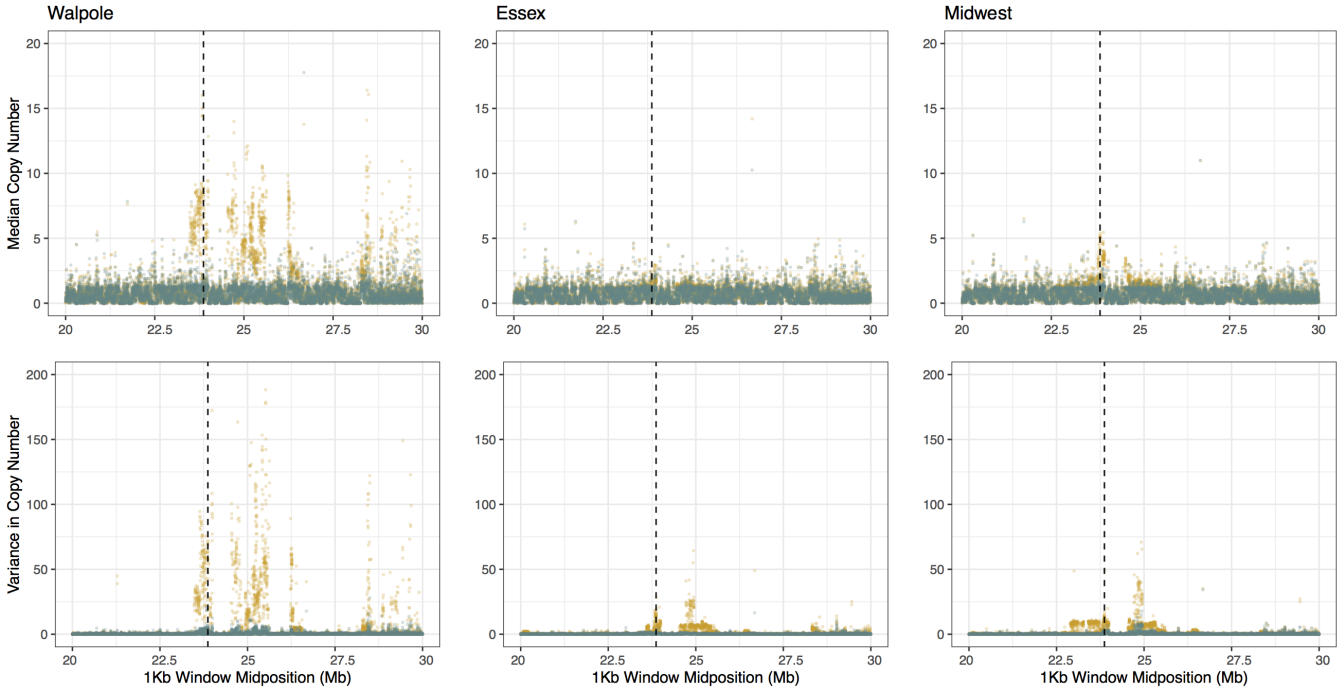
Sup Table 3. GO enrichment. Based on the 1% outliers of windows with excess differentiation between Walpole and Natural populations for a given level of Walpole diversity.

GO biological process complete	expected	Fold Enrichment	raw P value	FDR
RNA processing	5.51	3.27	1.45E-05	7.16E-03
RNA metabolic process	9.47	2.75	3.79E-06	2.24E-03
nucleic acid metabolic process	12.59	2.94	4.22E-09	1.25E-05
nucleobase-containing compound metabolic process	16.11	2.67	3.12E-09	1.84E-05
organic cyclic compound metabolic process	20.33	2.31	5.01E-08	5.93E-05
organic substance metabolic process	59.14	1.49	1.52E-05	6.90E-03
metabolic process	67.16	1.4	7.96E-05	3.14E-02
cellular nitrogen compound metabolic process	22.25	2.16	3.55E-07	3.51E-04
nitrogen compound metabolic process	45.92	1.65	1.71E-06	1.44E-03
cellular metabolic process	56.36	1.54	3.61E-06	2.37E-03
cellular process	81.17	1.34	7.54E-05	3.19E-02
heterocycle metabolic process	18.32	2.46	1.84E-08	2.72E-05
cellular aromatic compound metabolic process	19.48	2.41	1.24E-08	2.44E-05
primary metabolic process	53.96	1.54	8.71E-06	4.68E-03
macromolecule metabolic process	40	1.72	1.82E-06	1.34E-03
gene expression	11.38	2.28	9.08E-05	3.36E-02

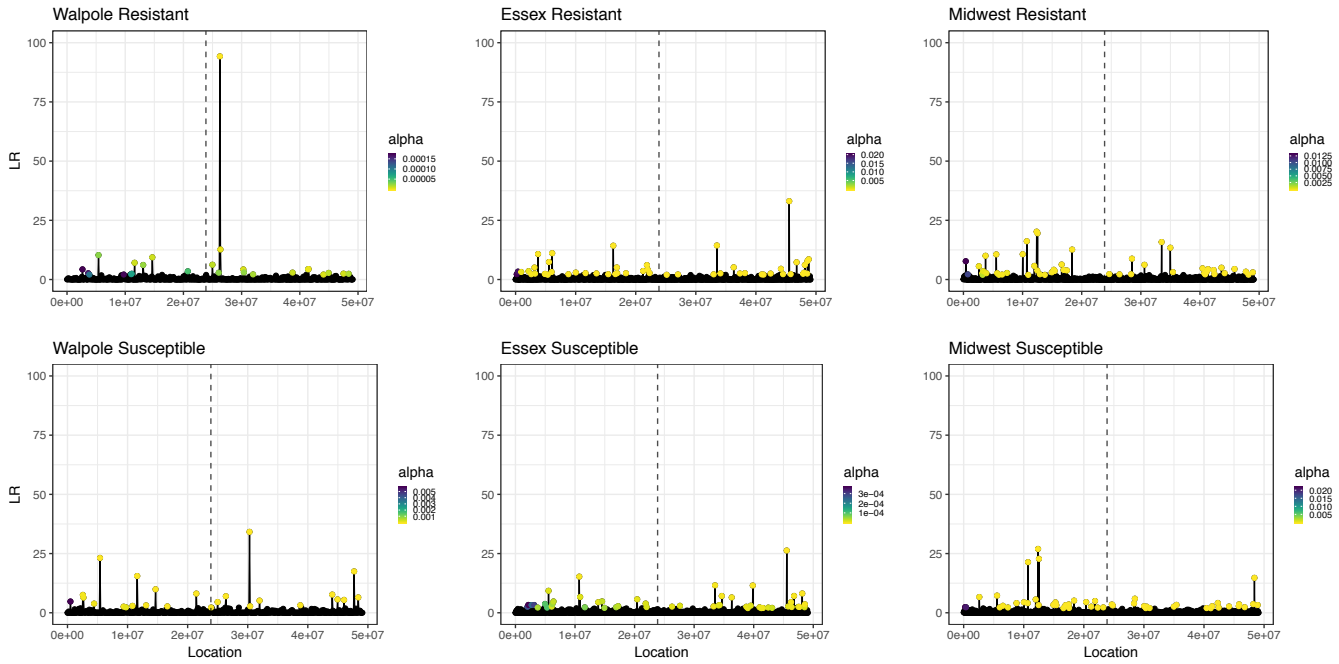


Sup Figure 1. The relationship between absolute diversity (D_{XT}), relative diversity (F_{ST}), and within population diversities (π) for among geographic region comparisons. For each comparison, points refer to the mean of 100 kb windows, with corresponding boxplots and density curves for each summary statistic.

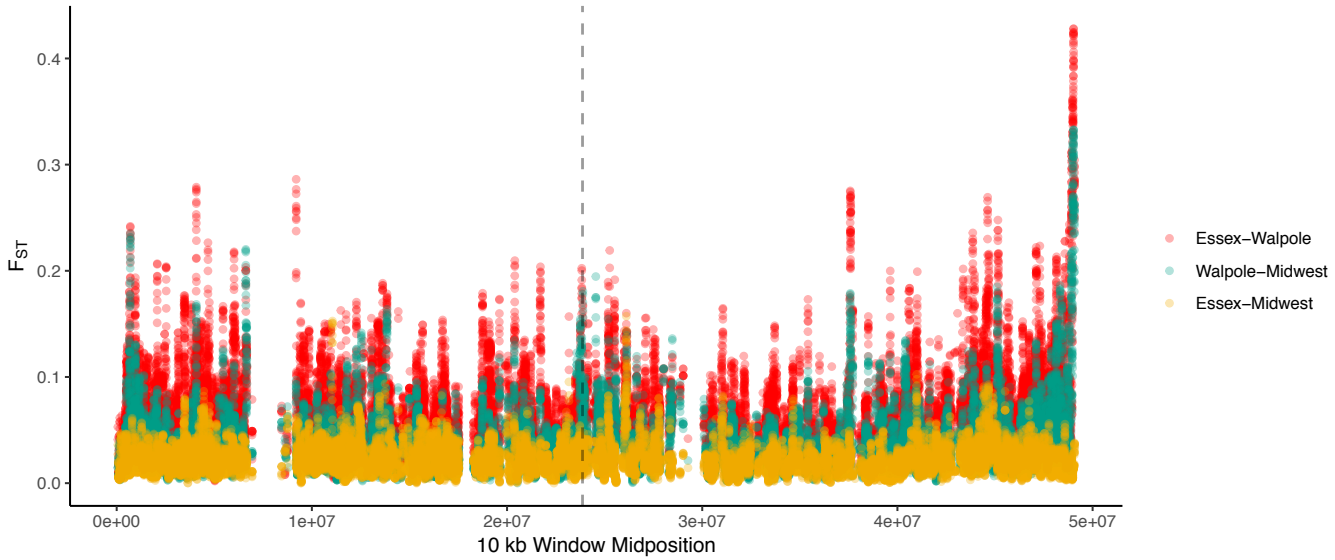
Sup Figure 2. Walpole diversity against differentiation between Walpole and Natural populations for 10 kb windows across the genome.



Sup Figure 3. Median and variance in *EPSPS* copy number for resistant (yellow) and susceptible (green) individuals across the chromosome 5 in each region. Dashed vertical lines indicate location of *EPSPS*.



Sup Figure 4. Sweepfinder2 likelihood ratio scores of a sweep occurring across chromosome 5. Scores were controlled for recombination rate variation and the genome-wide neutral site frequency spectrum. Alpha refers to the relative strength of recombination vs selection, with small values indicating strong selection (31).



Sup Figure 5. Relative differentiation (F_{ST}) for 10 kb windows among agricultural populations across chromosome 5. *EPSPS* delimited by the vertical dashed line.