Rapid adaptation and range expansion in response to agriculture over the last two centuries

Short title: Weed adaptation to modern agriculture

Julia Kreiner*[1,2], Sergio M. Latorre[3,4], Hernán A. Burbano[3,4], John R. Stinchcombe[5], Sarah P. Otto[2,6], Detlef Weigel[4], & Stephen I. Wright[5]

Affiliations:

[1]*Department of Botany, University of British Columbia, Vancouver, Canada;*
[2]*Biodiversity Research Centre, University of British Columbia, Vancouver, Canada;*
[3]*Centre for Life's Origins and Evolution, Department of Genetics, Evolution and Environment, University College London, London, UK;*
[4]*Department of Molecular Biology, Max Planck Institute for Biology Tübingen, Tübingen, Germany;*
[5]*Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, Canada;*
[6]*Department of Zoology, University of British Columbia, Vancouver, Canada;*

*Corresponding author email: julia.kreiner@ubc.ca

Abstract

North America has seen a massive increase in cropland use since 1800, accompanied more recently by the intensification of agricultural practices. Through genome analysis of present-day and historical samples spanning environments over the last two centuries, we studied the impact of these changes in farming on the extent and tempo of evolution in the native common waterhemp (*Amaranthus tuberculatus),* a now pervasive agricultural weed. Modern agriculture has imposed strengths of selection rarely observed in the wild (0.027-0.10), with striking shifts in allele frequency trajectories since agricultural intensification in the 1960s. An evolutionary response to this extreme selection was facilitated by a concurrent human-mediated range shift. By reshaping genome-wide diversity and variation for fitness, agriculture has driven the success of this 21st-century weed.

One Sentence Summary

Modern agriculture has dramatically shaped the evolution of a native plant into an agricultural weed through imposing strengths of selection rarely observed in the wild.

Main text

Agricultural practices across North America have rapidly intensified over the last two centuries, through cropland expansion (*1*), habitat homogenization (*2*), and increased chemical inputs (*3*). Since the beginning of the 1800s, cropland usage has expanded from 8 million to 200 million hectares in Canada and the United States alone (*1*). Since the 1960's, increased reliance on pesticides, irrigation, large-scale mechanization, and newly developed crop varieties have greatly improved the efficiency of food production in all farming sectors, a transformation oft-referenced as the agricultural "Green Revolution" (*4*, *5*). For pesticides, however, their effectiveness has been limited by the evolution of resistance across numerous pest species (*6*). While technological innovation for efficient food production has risen with increasing global food demands, the concomitant conversion of our landscape has become one of the foremost drivers of global biodiversity loss (*7*).

Species that have managed to survive, and even thrive, in the face of such extreme environmental change provide remarkable examples of rapid adaptation on contemporary timescales and illustrate the evolutionary consequences of anthropogenic impacts. One such species is common waterhemp (*Amaranthus tuberculatus*), which is native to North America and persists in large part in natural, riparian habitats (*8*), providing a unique opportunity to investigate the timescale and extent of contemporary agricultural adaptation in this prevalent weed. The genetic changes underlying weediness is particularly important to understand in *A. tuberculatus,* as it has become one of the most problematic weeds in North America due to widespread adaptation to herbicides, persistence in fields across seasons, and strong competitive ability with both soy and corn (*9*, *10*).

To understand how changing agricultural practices have shaped the success of a ubiquitous weed, we analyzed genomic data from contemporary paired natural and agricultural populations alongside historical samples collected from 1828 until 2018 (**Fig 1)**. With this design, we identify agriculturally adaptive alleles—those that are consistently higher in frequency in agricultural than in geographically close natural sites which constitute contrasts in selective pressures; track their frequency across nearly two centuries, and link the tempo of weed adaptation to demographic changes and key cultural shifts in modern agriculture.

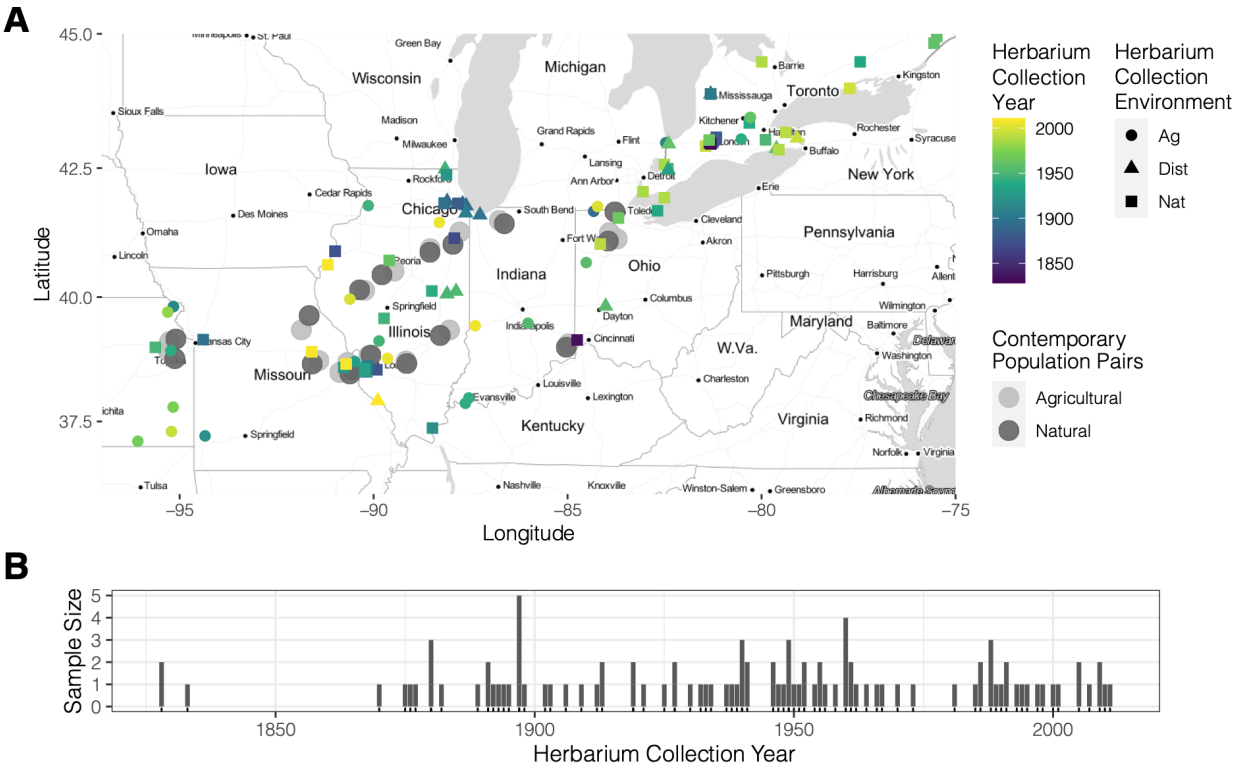

Fig 1. Sequenced waterhemp collections through space and time. **A)** Map of 14 contemporary paired natural-agricultural populations (n=187, collected and sequenced in Kreiner et al., 2021), along with 108 novel sequenced herbarium specimens dating back to 1828 collected across three environment types (Ag=Agricultural, Nat=Natural, Dist=Disturbed). **B)** Distribution of sequenced herbarium samples through time.

The genome-wide signatures of agricultural adaptation

To find alleles favored under current farming practices, we looked for those alleles that were consistently overrepresented in extant agricultural populations compared to neighboring natural populations (*11*), using Cochran–Mantel–Haenszel (CMH) tests (**Fig 2A**). Alleles involved in agricultural adaptation (the 0.1% of SNPs with lowest CMH p-values; n=2,055) are significantly enriched for 21 GO-biological process terms related to growth and development, cellular metabolic processes, and responses to biotic, external, and endogenous stimuli, including response to chemicals (**Table S1**). The importance of chemical inputs in shaping weed agricultural adaptation is clear in that the most significant agriculturally associated SNP (raw p-value = 8.551×10^{-11}, [FDR corrected] q-value = 0.00062) falls just 80 kb outside the gene protoporphyrinogen oxidase (*PPO*)—the target of PPO-inhibiting herbicides (**Fig 2B**). Other genes with the strongest agricultural associations include *ACO1*, which has been shown to confer oxidative stress tolerance (*12*); *HB13*, involved in pollen viability (*13*) as well as drought and salt tolerance (*13*); *PME3*, involved in growth via germination timing (*14*); *CAM1,* a regulator of senescence in response to stress (*15*, *16*); and both *CRY2* and *CPD*, two key regulators of photomorphogenesis and flowering via brassinosteroid signaling (*17–20*)

(**Table S2)**. These signals of agricultural adaptation are notable given that genome-wide differentiation among environments as measured by F_{ST} is negligible, a mere 0.0008 (with even lower mean F_{ST} between paired sites = -0.0029; **Fig 2C)**, suggesting that despite near panmixia among environments, strong antagonistic selection acts to maintain spatial differentiation for particular alleles.

To further investigate the extent to which herbicides shape adaptation to agriculture, we assayed signals of selection for two complex resistance variants—a deletion of codon 210 within *PPO,* which is causal for resistance to PPO-inhibiting herbicides (*21*), and amplification of 5-enolpyruvylshikimate-3-phosphate synthase (*EPSPS*), which confers resistance to glyphosate herbicides (*22*). Natural-vs-agricultural F_{ST} (highly correlated with the CMH test statistic, **Fig S1)** at the PPO210 deletion, 0.21, is higher than anywhere else in the genome and is even stronger when calculated within-population pairs (F_{ST}= 0.27) (**Fig 2C**). Similarly, the EPSPS amplification is ranked 20th among genome-wide biallelic F_{ST} values, 0.14 (within-pair F_{ST}= 0.22), in support of herbicides as a foremost driver of agricultural adaptation.

How differences in selection relative to migration among environments may mediate agricultural adaptation is pivotal for understanding the consequences of agricultural selection in natural environments and the persistence of resistance mutations through time. For 6 common alleles, previously shown to be causal for conferring herbicide resistance (*10*), as well as the top 30 independent CMH outliers, we implemented a Wright-Fisher allele-frequency-based migration-selection balance model to infer the relative strength of selection favoring resistance alleles in agricultural environments versus selection favoring susceptible alleles in natural environments at equilibrium. While resistance alleles were at intermediate frequency in agricultural populations, ranging from 0.08 to 0.35, they were rarer in natural populations, with frequencies from 0.04 to 0.22, consistent with on-going migration from agricultural into natural environments balanced by selection against these alleles in the absence of herbicides (**Fig 2D)**. Assuming these sites are at equilibrium, we inferred that the costs of resistance per migrant arriving in natural environments are stronger than the costs of susceptibility per migrant arriving in agricultural environments (per migrant costs: benefit ratio ranges from 1.09 for ALS653 to 4.2 for the PPO210 deletion, with a mean = 1.99; **Fig 2D, Table S3**). For the top 30 independent CMH outliers, the costs in natural environments were about equally likely to be stronger or weaker (12/28, 42%) than the benefits in agricultural environments, scaled to the migration rate (**Fig S2).** Thus, while substantial gene flow between agricultural and natural sites repeatedly introduces locally unfit alleles across environments, the spread of herbicide resistance alleles appears to be strongly constrained by their cost in herbicide-free, natural environments.

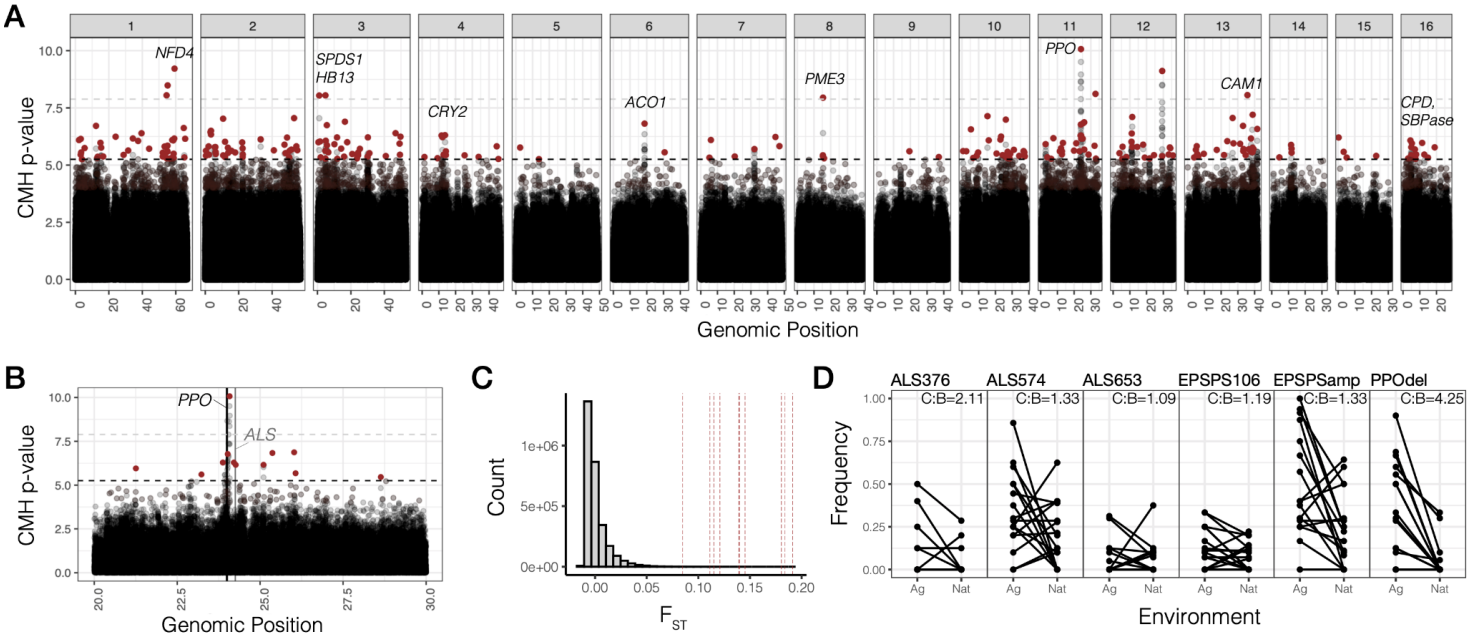

Fig 2. Signals of contemporary agricultural adaptation, gene flow, and antagonistic selection across the genome in *A. tuberculatus*. A) Results from Cochran–Mantel–Haenszel (CMH) tests for SNPs with consistent differentiation among environments across contemporary natural-agricultural population pairs. A 10% FDR threshold is indicated by the lower dashed horizontal black line, while the Bonferroni q-value < 0.1 cutoff is shown by the upper dashed horizontal gray line. Red points indicate focal adaptive SNPs after aggregating linked variation ($r^2 > 0.25$ within 1 Mb). Candidate agriculturally adaptive genes for peaks that are significant at a 10% FDR threshold shown. **B)** CHM results from the scaffold containing the most signficant CMH p-value, corresponding to variants linked to the PPO210 deletion conferring herbicide resistance and to the nearby herbicide-targeted gene *ALS*. **C)** Distribution of F_{ST} values between all agricultural and natural samples for ~3 million genome-wide SNPs (minor allele frequency > 0.05). Vertical lines indicate F_{ST} values for the 10 candidate genes named in A. **D)** Pairwise frequency of six common herbicide resistance alleles across agricultural and natural habitats sampled in 2018; the first four are nonsynonymous variants in *ALS* and *EPSPS*, the EPSPSamp is a 10 Mb-scale amplification that includes *EPSPS*, and the last one is an in-frame single-codon deletion in PPO (each dot represents on average ~5 individuals). Per migrant natural cost: agricultural benefit ratio relative to migration (C:B) is shown in the top right corner of each locus-specific comparison of frequencies across population pairs.

Agriculturally-adaptive alleles change rapidly with intensified regimes

With a strong set of agriculture-associated alleles (251 loci after aggregating linked SNPs), we searched for signatures of temporal evolution using newly collected whole genome sequence data from a set of historical samples (n=108) dating back to 1828, collected from natural, agricultural, and disturbed environments **(Fig 1)**. Of the 165 loci for which we had sufficient information in the historical SNP set (sequenced to 10x coverage on average), 151 were segregating with the same reference/alternate allele combination, and only three were invariant. To model allele frequency change through

time on these loci, we implemented logistic regressions of genotypes (within individual allele frequencies) at each locus on collection year, where the slope of the logit-transform is equivalent to the strength of selection (s). Because our historical collections sampled both natural and human-mediated environments through time, we were able to compare allele frequency trajectories and selection across environments.

Consistent with the rapid change in land use and farming practices in the recent past, the frequency of these 154 contemporary agricultural alleles has increased substantially over the last two centuries. Whereas in present-day natural environments agriculturally-adaptive alleles have increased by 6% on average since 1870, the earliest time point at which we have collections across environment types, these same alleles have increased by 22% in disturbed and agricultural environments (**Fig 3A**). This observed change greatly exceeds the expected change over this time period, under null processes (drift, migration, and selection) (null 95% interquantile range for allele frequency change in agricultural and disturbed sites = [3.3, 7.9%]; for change in natural sites = [-2.7, 2.0.%]). We generated these null expectations by randomly sampling the same number of observed loci across the genome and calculating their allele frequency change through time, where each of the 1000 randomized sets matched the frequency distribution observed for extant agricultural alleles (**Fig S4**) and were constrained to alleles across the genome that were at a higher frequency in agricultural compared to natural habitats. These randomizations were performed separately across environment types.

The considerable increase in frequency of these alleles across environments corresponds to remarkably strong selection even when estimated over century-long time periods. The 154 agriculture-associated alleles collectively exhibit $\tilde{s} = 0.011$ since the 1870s in agricultural and disturbed habitats but exhibit much weaker selection, $\tilde{s} = 0.0028$, in natural habitats (agricultural and disturbed null interquantile range = [0.0013, 0.0034]; natural null interquantile range = [-0.0009, 0.0009]). The range of selection estimated across loci varies between -0.098 and 0.075 in natural habitats, and -0.045 and 0.186 in agricultural and disturbed habitats (**Fig 3B, Fig S5)**. The top 15 agriculture-associated alleles that have experienced the strongest, significant selection over the last ~150 years include SNPs that map near *PPO, ACO1*, *CCB2, WRKY13, BPL3,* and *ATPD* (**Table S3)**. We find that both the total frequency change of agriculture-associated alleles and the estimated strength of selection in agricultural and disturbed environments are positively correlated with the extent of contemporary linkage disequilibrium around these loci (the number of SNPs with $r^2 > 0.25$ within 1Mb) (frequency change: $F = 5.16$, $p = 0.024$; strength of selection: $F = 3.99$, $p = 0.048$; **Fig S6**), consistent with theoretical expectations for the genomic signatures around alleles that have recently been impacted by positive selection (*23*, *24*).

Along with evidence of much stronger selection and frequency change of agriculturally adaptive alleles in agricultural versus natural environments over the last 150 years, we find that the trajectory of these alleles among environments varies considerably through time (**Fig 3C, Fig S7)**. While extant pairs of agricultural and natural populations are differentiated by 18% at these loci, this decreases as we look further back in time, so that around 1900, these alleles still had equal frequencies in both environments (predicted 1900 frequency in agricultural and disturbed sites = 41.9% [SE=2.7%], predicted 1900 frequency in natural sites = 38.6% [SE=2.7%]) (**Fig 3D)**. Moreover, when we split out

samples into those that predate or are subsequent to the intensification of agriculture during the Green Revolution, we find that the increase in frequency of agricultural alleles was negligible in agricultural and disturbed environments before the 1960s (predicted 1870-1960 change = 0.005), with the subsequent change near completely accounting for the observed rise in frequency of the alleles more common today in agricultural environments (predicted 1960-2018 change = 0.219, versus total 2018-1870 change = 0.221) (**Fig 3C**). Corresponding estimates of selection by logistic regression using only data from before 1960 shows no evidence of selection on these loci in disturbed and agricultural ($\tilde{s}$ = 0.0008, null interquantile range = [-0.0022,0.0010]) or in natural habitats ($\tilde{s}$= 0.0003, null interquantile range = [-0.002,0.002]). However, samples collected subsequent to 1960 reflect a dramatic shift in selection—a collective $\tilde{s}$= 0.027 in disturbed and agricultural environments and a collective $\tilde{s}$= 0.014 in natural environments (ag null interquantile range =[0.0032,0.0098]); nat null interquantile range =[-0.0028,0.0027]) (**Fig 3C; Fig S8)**. Together, these results suggest that while most contemporary agricultural alleles were present in historical populations, that these alleles only became associated with agricultural and human-managed sites over the last century, on timescales and rates consistent with the rapid uptake and intensification of agrochemicals, controlled irrigation, and mechanization in agriculture.

The historical trajectory of known herbicide resistance alleles epitomizes extreme selection over the last 50 years (**Fig 3D)**. Five out of seven known herbicide resistance loci present in our contemporary collection are absent from our historical samples, consistent with the suggested importance of resistance adaptation from *de novo* mutation (*25*, *26*). Only three out of 108 historical samples show variation for herbicide resistance, two samples homozygous for resistance at ALS574 and one heterozygous for resistance at ALS122—all of which were sampled after the onset of herbicide applications in the 1960s (**Fig 3D)**. Since 1960, we find that these seven known resistance alleles in our contemporary samples have collectively experienced selection of $\tilde{s}$= 0.099 (Z = 2.11, p = 0.035) per year—ranging from s > 0.097 for PPO210, s > 0.057 for EPSPS106, and s > 0.044 for ALS574 to no evidence of selection on ALS122 and ALS197 **(Fig 3D; Table S3**). As expected, selection has been particularly strong on these alleles in disturbed and agricultural environments ($\tilde{s}$= 0.1003, Z = 2.121, p = 0.034), but selection on these known resistance alleles remains high when estimated from samples taken over time in natural environments ($\tilde{s}$= 0.071, Z = 1.912, p = 0.056), where presumably the alleles have been recurrently introduced by migration from agricultural sites.

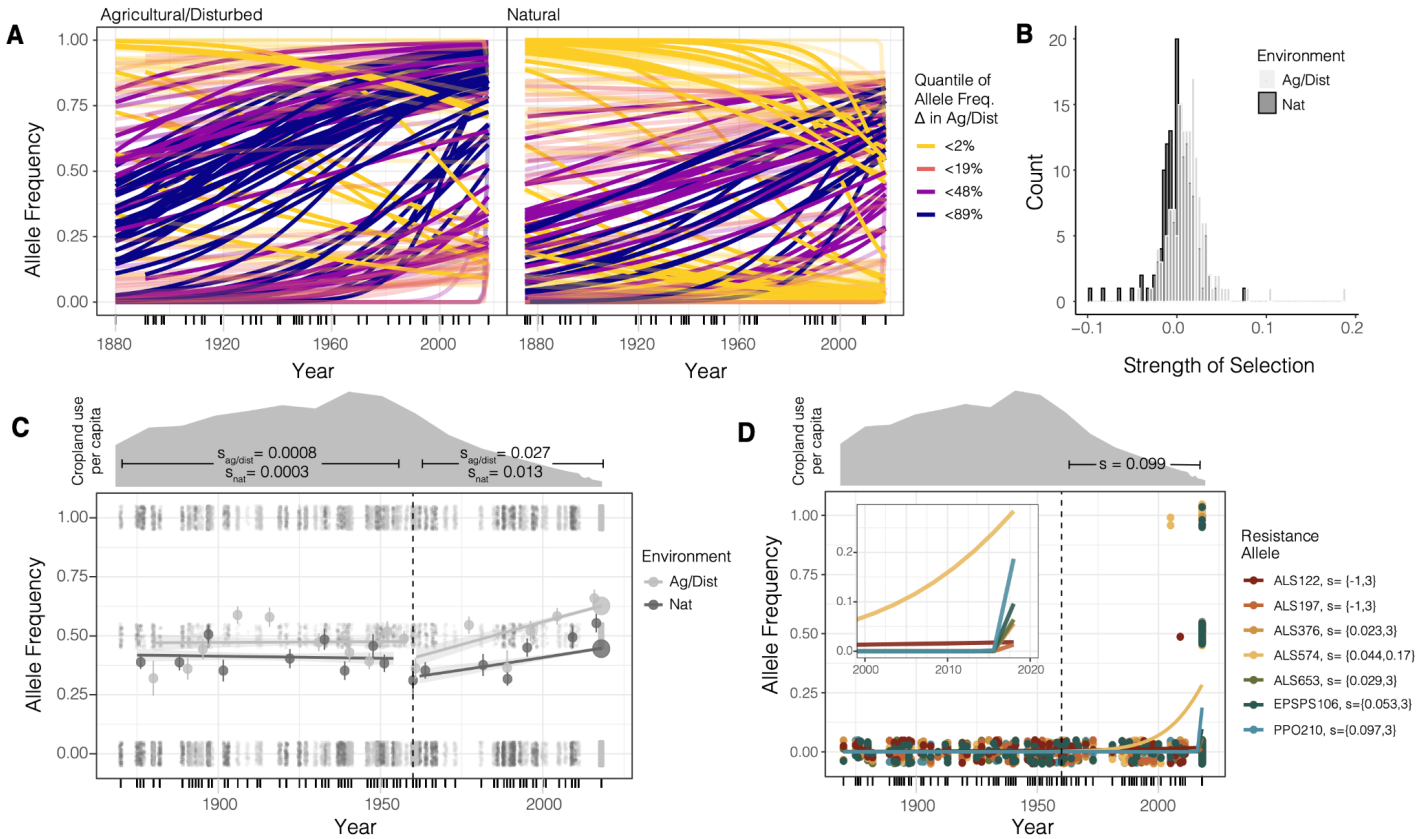

Fig 3. Genomic signatures of agricultural adaptation through time. A) Agricultural allele frequency trajectories for each locus, in agricultural and disturbed habitats (left), and natural habitats (right). Trajectories coloured by the quantile of frequency change in agricultural and disturbed habitats. Transparent lines indicate those with non-significant evidence of selection at α=0.05 after FDR=10% correction. **B)** The strength of selection on agricultural alleles for each locus in natural (dark gray) and agricultural and disturbed (light gray) habitats between 1870 and 2018. **C)** Agricultural allele frequency trajectories in each environment type, before and after the start of agricultural intensification in 1960. Vertical dashed line represents an inferred breakpoint in the data in a segmented regression. Environmental regression lines represent logistic fits to data that either predate or are subsequent to 1960. Large circles represent moving averages (over both loci and individuals) of allele frequencies, whereas dots represent raw genotype data for each locus and sample from which the allele frequency trajectory is estimated. Cropland use per capita in North America data from (*1*)**,** rescaled by use in 1600. **D)** The trajectory of alleles at known herbicide resistance loci through time, fit by logistic regression for each of the seven alleles present in our contemporary data. Dots represent genotypes for each historical and contemporary sample at each herbicide resistance locus. 95% credible interval of the maximum likelihood estimate of selection between 1960-2018 provided in the legend for each resistance allele.

Concurrent temporal shifts in ancestry underlie agricultural adaptation

Finally, we explored whether historical demographic change over the last two centuries has played a role in agricultural adaptation. Early taxonomy described two different *A. tuberculatus* varieties as separate species, with few distinguishing characteristics (seed

dehiscence and tepal length (*8*)). Sauer's 1955 revision of the genus, which used herbarium specimens to gauge the distribution and migration of congeners over the last two centuries (*27*), led him to describe an expansion of the southwestern var. *rudis* type (at the time, *A. tamariscinus* (Sauer)) northeastward into the territory of var. *tuberculatus (A. tuberculatus* (Sauer)), sometime between 1856-1905 and 1906-1955. Our sequencing of over 100 herbarium samples dating back to 1828, combined with nearly 200 contemporary sequences, allowed us to directly observe the change in the distribution of these two ancestral types, adding resolution to Sauer's morphological observations of the species' contemporary range shifts at a genome-wide level and over more recent timescales.

Range-wide, we see clear shifts in the distribution of var. *rudis* ancestry based on faststructure inference at K=2 (**Fig S9)** across three-time spans, 1830-1920, 1920-1980, and 1980-2018 (timespan: $F = 5.47$, $p = 0.0045$), and particularly so in the East (timespan x longitude: $F = 5.49$, $p = 0.0045$), consistent with a recent expansion of var. *rudis* ancestry (**Fig 4A)**. Furthermore, we see strong state and province-specific shifts in ancestry through time in our historical sequences (time span by state interaction: $F = 4.22$, $p = 7 \times 10^{-5}$), highlighting not only the shift of var. *rudis* eastwards (with increases through time in Ontario, Ohio, Illinois, and Missouri) but also the very recent introduction of var. *tuberculatus* ancestry into the most eastwards part of the range in Kansas (**Fig 4B)**. *A. tuberculatus* demography thus appears to have been drastically influenced by human-mediated landscape change over the last two centuries, consistent with the massive recent expansion of effective population size we have previously inferred over this same timeframe (*26*). That this shift has been most notable over the last 40 years is further consistent with the timescale of rampant herbicide resistance evolution within the species (*10*, *26*, *28*), suggesting selection on resistance may facilitate the colonization of var. *rudis* ancestry outside its historical range. Along these lines, we find this contemporary expansion has facilitated the sorting of var. *rudis* ancestry across environments (a longitude by time span by environment interaction: $F = 5.13$, $p = 4\ 10^{-5}$; **Fig 4C**), with increasing overrepresentation of var. *rudis* ancestry in agricultural and disturbed environments in the eastern portion of the range through time, as previously suggested (*11*).

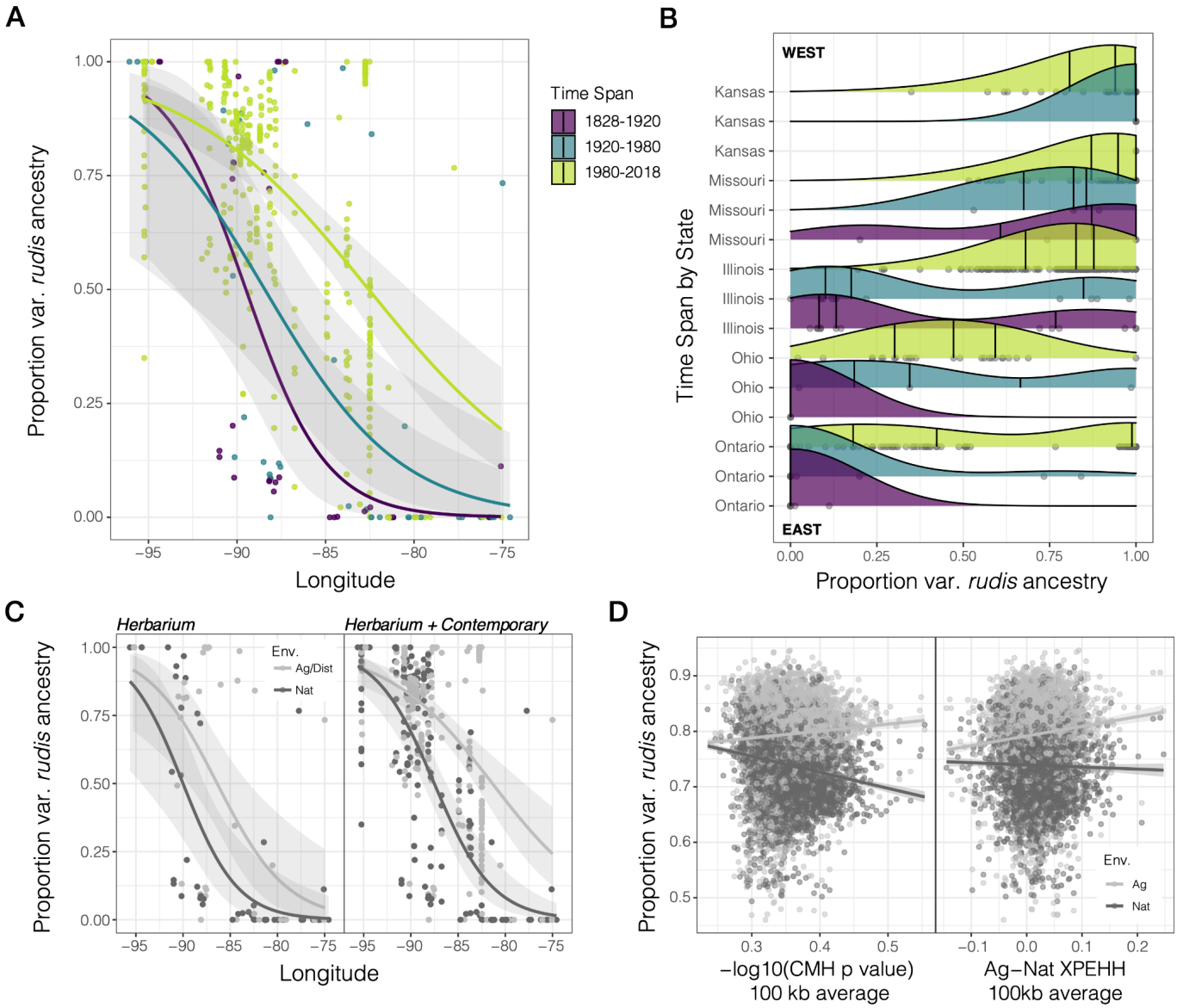

Fig 4. Temporal shifts in the distribution of var. *rudis* ancestry have facilitated polygenic agricultural adaptation. A) Longitudinal clines in individual-level var. *rudis* ancestry over three timespans, illustrating the expansion of var. *rudis* ancestry eastwards over the last two centuries. **B)** The distribution of individual-level var. *rudis* ancestry by state and through time, illustrating state-specific changes in ancestry. Vertical lines represent first, second, and third quantiles of ancestry within each timespan and state. Timespans indicated in A) **C)** Increasing sorting of individual-level var. *rudis* ancestry into agricultural environments on contemporary timescales. **D)** Environment-specific metrics of selection (CMH p-value and cross-population extended haplotype homozygosity (XPEHH)) across the genome in 100 kb windows positively correlate with var. *rudis* ancestry in agricultural, but not natural habitats.

To investigate whether agricultural adaptation has preferentially favored var. *rudis* ancestry, we reconstructed fine-scale ancestry across the genome. Based on analyses in 100 kb windows, we find a least squares mean of 5.6% more var. *rudis* ancestry genome-

wide in agricultural environments compared to the adjacent natural habitat (**Fig S10)**. The environment-specific proportion of var. *rudis* ancestry is not only positively correlated with recombination rate (F = 16.67, p = 4.5 x 10^{-5}) and gene density (F = 5.85, p = 0.016) but also with SNP and haplotype-based evidence of environment-specific selection. Agricultural, but not natural populations, have an excess of cross-population haplotype homozygosity (agricultural vs natural XPEHH) and within-pair environmental differentiation (CMH p-value) in genomic regions of high var. *rudis* ancestry (Env x XPEHH: F=9.34, p=0.002; Env x CMH: F=99.70, $p < 10^{-16}$; **Fig 4D**), implying that ancestry composition genome-wide in large part determines the extent of polygenic agricultural adaptation. Together, these findings suggest that the expansion of var. *rudis* ancestry across the range, particularly in the last 40 years, has facilitated adaptation to novel agricultural selective pressures through providing preadapted genetic variation.

In summary, agricultural adaptation in *A. tuberculatus*, a native plant in North America, has occurred over extremely rapid timescales, facilitated by range shifts in response to the agriculturalization of its native habitat. The human-mediated expansion of the southern lineage of the species northeastwards since the later half of the 20th century has introduced new genetic variation across the genome on which selection in agricultural settings could act. Through our paired sampling design, we identified 251 independent SNPs across 240 genes that are implicated in agricultural adaptation; these genes tend to be enriched for expanded southwestern ancestry, with functions affecting growth, development, abiotic tolerance, and herbicide resistance. Concurrent with the intensification of agriculture, the prevalence of agricultural alleles has increased rapidly over just the last 60 years, in agricultural environments by nearly 3% per year, and even in natural sites by more than 1% per year. The first empirical estimates of selection coefficients for herbicide resistance provided here—10% per year range-wide over a 60 year period—emphasizes the long lasting impact of selection on genetic variation even across heterogeneous environments. Modern, industrial agriculture thus imposes strengths of selection rarely observed in the wild.

These results highlight that anthropogenic change not only leads to the formation of new habitats but also provides an opportunity for range expansion that may facilitate and feedback with local adaptation, reshaping genetic variation for fitness within native species.

References and Notes

1. K. Klein Goldewijk, A. Beusen, J. Doelman, E. Stehfest, Anthropogenic land use estimates for the Holocene – HYDE 3.2. *Earth Syst. Sci. Data*. **9**, 927–953 (2017).

2. T. G. Benton, J. A. Vickery, J. D. Wilson, Farmland biodiversity: is habitat heterogeneity the key? *Trends in Ecology & Evolution*. **18** (2003), pp. 182–188.

3. D. Kleijn, F. Kohler, A. Báldi, P. Batáry, E. D. Concepción, Y. Clough, M. Díaz, D. Gabriel, A. Holzschuh, E. Knop, A. Kovács, E. J. P. Marshall, T. Tscharntke, J. Verhulst, On the relationship between farmland biodiversity and land-use intensity in Europe. *Proc. Biol. Sci.* **276**, 903–909 (2009).

4. C. Mann, Reseeding the Green Revolution. *Science* (1997), doi:10.1126/science.277.5329.1038.

5. P. L. Pingali, Green revolution: impacts, limits, and the path ahead. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 12302–12308 (2012).

6. D. Pimentel, H. Lehman, *The Pesticide Question: Environment, Economics and Ethics* (Springer Science & Business Media, 1993).

7. F. Zabel, R. Delzeit, J. M. Schneider, R. Seppelt, W. Mauser, T. Václavík, Global impacts of future cropland expansion and intensification on agricultural markets and biodiversity. *Nat. Commun.* **10**, 2844 (2019).

8. J. Sauer, REVISION OF THE DIOECIOUS AMARANTHS. *Madroño*. **13**, 5–46 (1955).

9. 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).

10. P. J. Tranel, Herbicide resistance in Amaranthus tuberculatus†. *Pest Manag. Sci.* **77**, 43–54 (2021).

11. J. M. Kreiner, A. Caballero, S. I. Wright, J. R. Stinchcombe, Selective ancestral sorting and de novo evolution in the agricultural invasion of Amaranthus tuberculatus. *Evolution* (2021), doi:10.1111/evo.14404.

12. W. Moeder, O. Del Pozo, D. A. Navarre, G. B. Martin, D. F. Klessig, Aconitase plays a role in regulating resistance to oxidative stress and cell death in Arabidopsis and Nicotiana benthamiana. *Plant Mol. Biol.* **63**, 273–287 (2007).

13. J. V. Cabello, R. L. Chan, The homologous homeodomain-leucine zipper transcription factors HaHB1 and AtHB13 confer tolerance to drought and salinity stresses via the induction of proteins that stabilize membranes. *Plant Biotechnol. J.* **10**, 815–825 (2012).

14. S. Guénin, J. Hardouin, F. Paynel, K. Müller, G. Mongelard, A. Driouich, P. Lerouge, A. R. Kermode, A. Lehner, J.-C. Mollet, J. Pelloux, L. Gutierrez, A. Mareck, AtPME3, a ubiquitous cell wall pectin methylesterase of Arabidopsis thaliana, alters the metabolism of cruciferin seed storage proteins during post-germinative growth of seedlings. *J. Exp. Bot.* **68**, 1083–1095 (2017).

15. S. Zhou, L. Jia, H. Chu, D. Wu, X. Peng, X. Liu, J. Zhang, J. Zhao, K. Chen, L. Zhao, Arabidopsis CaM1 and CaM4 Promote Nitric Oxide Production and Salt Resistance by Inhibiting S-Nitrosoglutathione Reductase via Direct Binding. *PLoS Genet.* **12**, e1006255 (2016).

16. C. Dai, Y. Lee, I. C. Lee, H. G. Nam, J. M. Kwak, Calmodulin 1 Regulates Senescence and ABA Response in Arabidopsis. *Front. Plant Sci.* **9**, 803 (2018).

17. H. Guo, H. Yang, T. C. Mockler, C. Lin, Regulation of flowering time by Arabidopsis photoreceptors. *Science*. **279**, 1360–1363 (1998).

18. T. Mockler, H. Yang, X. Yu, D. Parikh, Y.-C. Cheng, S. Dolan, C. Lin, Regulation of photoperiodic flowering by Arabidopsis photoreceptors. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 2140–2145 (2003).

19. W. Wang, X. Lu, L. Li, H. Lian, Z. Mao, P. Xu, T. Guo, F. Xu, S. Du, X. Cao, S. Wang, H. Shen, H.-Q. Yang, Photoexcited CRYPTOCHROME1 Interacts with Dephosphorylated BES1 to Regulate Brassinosteroid Signaling and Photomorphogenesis in Arabidopsis. *Plant Cell*. **30**, 1989–2005 (2018).

20. J. Li, Y. Li, S. Chen, L. An, Involvement of brassinosteroid signals in the floral-induction network of Arabidopsis. *J. Exp. Bot.* **61**, 4221–4230 (2010).

21. F. E. Dayan, P. R. Daga, S. O. Duke, R. M. Lee, P. J. Tranel, R. J. Doerksen, Biochemical and structural consequences of a glycine deletion in the α-8 helix of protoporphyrinogen oxidase. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. **1804**, 1548–1556 (2010).

22. H. M. Cockerton, S. S. Kaundun, L. Nguyen, S. J. Hutchings, R. P. Dale, A. Howell, P. Neve, Fitness cost associated with enhanced EPSPS gene copy number and glyphosate resistance in an Amaranthus tuberculatus population. *Cold Spring Harbor Laboratory* (2021), p. 2021.01.09.426028.

23. M. Przeworski, The signature of positive selection at randomly chosen loci. *Genetics*. **160**, 1179–1189 (2002).

24. Y. Kim, R. Nielsen, Linkage disequilibrium as a signature of selective sweeps. *Genetics*. **167**, 1513–1524 (2004).

25. N. J. Hawkins, C. Bass, A. Dixon, P. Neve, The evolutionary origins of pesticide resistance. *Biol. Rev. Camb. Philos. Soc.* (2018), doi:10.1111/brv.12440.

26. J. M. Kreiner, G. Sandler, A. J. Stern, P. J. Tranel, D. Weigel, J. Stinchcombe, S. I. Wright, Repeated origins, widespread gene flow, and allelic interactions of target-site herbicide resistance mutations. *Elife*. **11**, e70242 (2022).

27. J. Sauer, Recent Migration and Evolution of the Dioecious Amaranths. *Evolution*. **11**, 11–31 (1957).

28. P. J. Tranel, C. W. Riggins, M. S. Bell, A. G. Hager, Herbicide resistances in Amaranthus tuberculatus: a call for new options. *J. Agric. Food Chem.* **59**, 5808–5812 (2011).

29. S. M. Latorre, P. L. M. Lang, H. A. Burbano, R. M. Gutaker, Isolation, Library Preparation, and Bioinformatic Analysis of Historical and Ancient Plant DNA. Curr Protoc Plant Biol. 5, e20121 (2020).

30. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 34 (2018), pp. i884–i890.

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

32. A. Peltzer, G. Jäger, A. Herbig, A. Seitz, C. Kniep, J. Krause, K. Nieselt, EAGER: efficient ancient genome reconstruction. Genome Biol. 17, 60 (2016).

33. A. Ginolhac, M. Rasmussen, M. T. P. Gilbert, E. Willerslev, L. Orlando, mapDamage: testing for damage patterns in ancient DNA sequences. Bioinformatics. 27, 2153–2155 (2011).

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

35. D. M. Emms, S. Kelly, OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16, 157 (2015).

36. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. J. Mol. Biol. 215, 403–410 (1990).

37. 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).

38. O. Delaneau, J.-F. Zagury, J. Marchini, Improved whole-chromosome phasing for disease and population genetic studies. Nat. Methods. 10, 5–6 (2013).

39. A. Raj, M. Stephens, J. K. Pritchard, fastSTRUCTURE: variational inference of population

structure in large SNP data sets. Genetics. 197, 573–589 (2014).

40. B. Pasaniuc, S. Sankararaman, G. Kimmel, E. Halperin, Inference of locus-specific ancestry in closely related populations. Bioinformatics. 25, i213–21 (2009).

Acknowledgements: We appreciate the pivotal contribution of numerous herbaria towards this research, especially the help of Eric Knox at Dean Herbarium at Indiana State University, Jamie Lynn Minnaert-Grote at the University of Illinois, Tedesse Mesfin at the University of Ohio Herbarium, Anton Reznicek at the University of Michigan Herbarium, Jim Solomon at the Missouri Botanical Gardens, Caleb Morse at the McGregor Herbarium at Kansas State University, Tyler Smith and Song Wang at Agricultural and Agrifood Canada, and Deb Metsger and Tim Dickinson at the Royal Ontario Museum. We thank the Whitlock lab (University of British Columbia), as well as Aneil Agrawal and Tyler Kent (University of Toronto) for input on the work; Christa Lanz and Rebecca Schwab (Max Planck Institute) for coordinating sequencing of herbarium samples; Ella Reiter (University of Leipzig) for scheduling and coordinating logistics for clean room facility work; and Patricia Lang, Sonja Kersten and Heike Budde (Max Planck Institute) for advice on molecular protocols troubleshooting.

Funding: JMK was supported by the Biodiversity Research Institute at the University of British Columbia and a Killam Fellowship. SIW was supported by a NSERC discovery grant and a Canada research chair. JRS was supported by a NSERC discovery grant. SML, HAB and DW were supported by the Max Planck Society.

Author Contributions: JMK, JRS, and SIW conceptualized the paired sampling design, JMK, HAB, DW, JRS, and SIW conceptualized the use of herbarium data, JMK performed contemporary collections and curated the herbarium samples, SML and HAB conceptualized and designed the molecular work with herbarium specimens, SML coordinated the clean room facility work, JMK and SML performed DNA extraction and library preparations of herbarium tissue, SML oversaw the sequencing of herbarium specimens. JMK performed analyses with input from SPO, SIW, and JRS. SPO wrote the migration-selection and maximum likelihood models. JMK wrote and revised the paper with inputs from all authors.

Competing interests: The authors declare that they have no competing interests.

Data and materials availability: All novel sequence data will be archived on SRA, while scripts and and accompanying metadata will be archived on Github and Dryad upon acceptance.

Supplementary Materials

Materials & Methods

Herbarium collections

In 2019, we obtained 10 mg tissue collections of herbarium specimens from 7 herbaria across Canada and the United States and one governmental organization: the Royal Ontario Museum Herbarium, the Museum of Biological Diversity at Ohio State University Herbarium, the Dean Herbarium at Indiana State University, the Michigan State University Herbarium, the Illinois Natural History Survey Herbarium, Missouri Botanical Gardens, The McGregor Herbarium at Kansas State University, and Agriculture and Agrifood Canada. We selected samples to have an even representation of habitats through time. Samples were classified as natural (n=54), agricultural (n=28), or disturbed (n=20) based on collectors' annotations on each plate: any reference to a cultivated field was treated as an 'agricultural' collection; general environmental descriptions such as dry grassland or riverbank was treated as a 'natural' collection; and reference to disturbed soil, railroad tracks, or manicured or managed land was treated as a 'disturbed' collection. For inference of contemporary allele frequency and ancestry change through time, samples collected from disturbed habitats were grouped together with the agricultural category—in both of which waterhemp exists as a weed (**Table S5)**. When geographic coordinates were not provided, we referred to the state, county, section, intersection, and landmark descriptions to infer the geographic coordinate of a given sample. In total, we collected samples from 172 specimens, 108 of which were selected for whole-genome sequencing.

Herbarium DNA extractions & library preparations

The work was performed in the ancient DNA lab at the University of Tübingen. For DNA extraction of the herbarium samples, we followed basic protocol 1 outlined in (*29*). Briefly, under sterile conditions, ~10 mg of each sample were ground and incubated with N-phenacylthiazolium bromide (PTB)-based mix overnight to lyse DNA. After a shredding step with QIAshredder spin columns, DNA was purified and eluted with DNAeasy Mini spin columns. Sequencing libraries were prepared using the basic protocol 2 outlined in (*29*), performing blunt-end repair, adapter ligation, a fill-in reaction, indexing, and finally PCR amplification (10 cycles) and a cleaning step. The libraries were sequenced on an Illumina NovaSeq instrument on a single flow cell. The sequencing run produced ~3,442 Gb data, an average of 32 Gb per sample.

Mapping, damage correction, SNP calling and filtering

We removed adapters, polyQ tails, and merged reads from herbarium sequencing reads using fastp (*30*). Because of the small fragment size of historical DNA, this resulted in a sizable loss of sequence coverage, from 46X coverage to a mean of 11X coverage. On

average, 89% of merged reads mapped to the female reference genome from (*31*), suggesting low rates of contamination by exogenous DNA. Finally, we performed de-duplication of merged reads with DeDup (*32*), which is optimized for merged paired-end sequencing data. This resulted in a final mean per-sample coverage of 9.7X.

We used the program MapDamage (*33*) to quantify damage patterns in the historical DNA. The fraction of C deamination, which leads to C-to-T substitutions, was low, at the first base ~2% on average across samples, barely inflated above the C-to-T substitution rate across the rest of the reads (**Fig S3**). Nonetheless, the fraction of C-to-T substitutions at the first base was positively correlated with the age of the samples (**Fig S3**). We thus used MapDamage to rescale mapping quality scores to take into account the patterns of DNA damage. We called SNPs with freebayes (v1.3.2) in 100 kb regions in parallel across the genome, merged, and then filtered SNPs based on quality (QUAL > 30) and missing data (< 0.30).

Herbicide resistance alleles in herbarium samples were identified based on known locations of non-synonymous substitutions within ALS and EPSPS. Initially, two genotype calls from herbarium samples that predated the onset of ALS herbicide use in the 1950s, showed standing variation for resistance at ALS574 and ALS122: one individual heterozygous for Trp-574-Leu collected in 1930 from a sandy agricultural field in St. Louis, Missouri, USA (HB0973); and another individual heterozygous for Ala-122-Ser collected in 1895 from a corn field in Fayette, Ohio, USA (HB0914). Upon further inspection, read-level support for resistance alleles was low with the allelic-bias at these genotype calls being highly skewed (reference to alternate ratio = 1:9 and 2:18, respectively). Similarly, one individual collected in 1967 from the Bottom of Maumee River, Ohio (HB0977) was heterozygous for ALS122, but the alternate resistance allele had support at only one read (reference to alternate ratio =1:7). We subsequently dropped these genotype calls from analyses of selection on herbicide resistance alleles through time.

Metrics of differentiation across Environments: CMH, F_{ST}, & XPEHH

We used the 7,262,599 genome-wide high-quality SNPs called from contemporary agricultural-natural paired populations (n=187 individuals total from 17 pairs of populations, 34 populations in total) from (*11*) (Fig 1). Previously, these data had been only used for genome-wide PCA and faststructure based individual-level ancestry estimates. To make use of our paired sampling design, we used plink (*34*) to perform a Cochran–Mantel–Haenszel test, testing an (environment by SNP | pair) effect after applying a minor allele frequency cutoff of 0.01. We identified candidate agriculturally-adaptive genes based on the nearest gene (bedtools closest) to each LD-clumped, FDR q-value < 0.1 SNP. We found the *Arabidopsis thaliana* orthologues of our *A. tuberculatus* genes with orthofinder (*35*). For genes where orthofinder found no *A. tuberculatus* orthologue and in which our annotation identified no orthologue in closely related species based on gene expression data, we used blastn (*36*) to perform a conclusive search for similar genes across species.

Additionally, we used plink to calculate Weir and Cockerham's F_{ST}, both between all natural and agricultural samples, and between environments within each population pair,

which we later averaged to obtain the mean pairwise F_{ST}. For calculation of Fst at the EPSPS amplification, we recoded individuals as 0, 1, 2 based on copy number amplitude (<1.5, 1.5 < copies < 2.5, and >2.5, respectively). We used selscan (*37*) to calculate the cross-population extended haplotype homozygosity, after read-back and population-level phasing with Shapeit2 (*38*), both of which required knowledge of recombination rates, which we supplied in the format of our imputed LD-based map from (*31*).

Models of Migration-Selection Balance

We modeled migration-selection balance between natural and agricultural habitats in our contemporary data for 6 common target-site resistance alleles, based on a two-patch, allele-focused model. In each patch, we modeled the frequency of the resistance allele and the susceptible (x_S, x_R and y_S, y_R, respectively) from year to year as:

$$x_S^* = x_S{}^2 W_{BB} + x_S x_R \quad W_{Bb}, \quad x_R^* = x_S x_R \quad W_{Bb} + x_R{}^2 W_{bb}$$

$$y_S^* = y_S{}^2 V_{BB} + y_S y_R \quad V_{Bb}, \quad y_R^* = y_S y_R \quad V_{Bb} + y_R{}^2 V_{bb}$$

where xS*, yS*represent the frequency of the susceptible allele and and xR*, yR*represent the frequency of the resistance allele after a bout of selection in natural (*x*) and agricultural (*y*) sites within a generation. We then allowed for migration of surviving genotypes and modeled their frequency in the next generation as follows:

$$x_S' = (1 - m_N)\left(\frac{x_S^*}{x_S^* + x_R^*}\right) + m_N\left(\frac{y_S^*}{y_S^* + y_R^*}\right),$$

$$x_R' = (1 - m_N)\left(\frac{x_R^*}{x_S^* + x_R^*}\right) + m_N\left(\frac{y_R^*}{y_S^* + y_R^*}\right)$$

$$y_S' = (1 - m_A)\left(\frac{y_S^*}{y_S^* + y_R^*}\right) + m_A\left(\frac{x_S^*}{x_S^* + x_R^*}\right),$$

$$y_R' = (1 - m_A)\left(\frac{y_R^*}{y_S^* + y_R^*}\right) + m_A\left(\frac{x_R^*}{x_S^* + x_R^*}\right)$$

where mN and mA represent immigration rates into natural and agricultural sites, respectively. Assuming additivity (*h*=0.5) and that migration at the loci is much weaker than selection (*m* << *s*), a given pair of populations is expected to approach a polymorphic equilibrium, where:

$\frac{s_A}{m_A} = \frac{2(y_S' - x_S')}{y_S' y_R'}$ in natural patches and $\frac{s_N}{m_N} = \frac{2(x_R' - y_R')}{x_S' x_R'}$ in agricultural patches

While it is not possible to solve for selection directly in the absence of data on migration rates, these formulae allow us to estimate the strength of divergence by inferring the strength of selection relative to migration in natural ($\frac{s_N}{m_N}$) and agricultural ($\frac{s_A}{m_A}$) environments, as presented in **Table S3**. The ratio of these metrics gives the ratio of the cost faced per migrant arriving in natural environments versus the benefit per migrant in

agricultural environments, assuming that the pair of populations is near equilibrium. We note that the approach to migration-selection balance occurs exponentially at a rate proportional to the selection coefficient (when $m << s << 1$) and so should occur rapidly at sites under strong selection (**Supplemental Index 1**).

Logistic models of temporal allele frequency change

We used CMH outliers from the contemporary paired population scan to investigate patterns of agricultural-allele frequency change over the last two centuries. We were interested in tracking independent allele frequency trajectories, so from the 403 SNPs with CMH p-values that exceeded 10% FDR correction ($p < 6 \times 10^{-6}$), we performed a subsequent clumping step, effectively retaining a set of largely unlinked SNPs (**Fig S11)** that represent the most significant SNP in a particular region. Specifically, we used plink --clump, to find the most significant hit genome-wide, scan 1 Mb around it, and exclude any SNP from the resulting output that is in $r^2 > 0.25$ with the focal SNP. This algorithm is repeated until all SNPs passing the genome-wide significance threshold have been clumped. This resulted in 251 loci that on average showed a 17.9% allele frequency difference between extant agricultural and natural environments. Because some of the alternate alleles across these loci were more frequent in natural environments, we redefined the alleles based on which one was more common in agricultural compared to natural sites.

We then found the intersection of these agriculture-associated alleles, identified in our contemporary paired collections, with the historical, filtered SNPs from the herbarium sequence data. 154 loci were present in the historical samples with the same reference/alternate allele combinations. We extracted a matrix of 0, 0.5, 1 values, representing the frequency of the agricultural allele for each locus within each individual, for samples from both our contemporary and historical collections. Combining these individual agricultural allele frequencies at each locus across historical and contemporary datasets, we then performed a logistic regression in R (glm function, family="binomial") of genotype on collection year, separately on samples from either natural or agricultural environments. From each logistic regression, we extracted the logit-transformed slope (selection coefficient, s), p-value, and standard error, as well as the predicted value (allele frequency) at 1870 and 2018, representing the minimum sample year and maximum sample year. While we have samples dating back to 1828, we constrained this analysis to samples collected after 1870, as the density of samples before then is low (n=4), with no representation of samples from agricultural environments.

The total allele frequency change at each locus was calculated by taking the difference between the predicted frequency of the allele in 2018 and 1870. We merged the output of these locus-specific logistic regressions in agricultural environments, with both SNP and haplotype-based statistics from these same individuals to identify contemporary correlates of the magnitude of allele frequency change and selection through time. Specifically, we examined how well contemporary recombination rate, XPEHH, the CMH p-value, the number of SNPs in linkage ($r^2 > 0.25$) with the focal SNP (< 1Mb; i.e. number of SNPs in a clump), and the distance between linked SNPs, explained both the total allele frequency change and the estimated strength of selection (**Fig S6)**.

We also performed a separate set of analyses, where a logistic regression was used to analyze the trajectory of all agricultural alleles or known herbicide resistance alleles at once, first across samples from natural environments and then for samples from agricultural and disturbed environments ('genotype ~ year + locus'; **Fig 3D**). We further partitioned samples in each environment to those that predate or are subsequent to the 1960s, to infer the importance of the intensification of agriculture and herbicides in shaping the strength of selection on contemporary agricultural loci. For each of the four logistic regressions ran on these partitioned sets of data, the slope of the year term represents a joint estimate of the strength of selection for agricultural alleles, between 1870-1960 or 1960-2018, in natural or agricultural environments. We refer to this joint estimate of selection at multiple loci as s.

To test whether a comparison of selection before and after the 1960's was statistically supported, we also compared our full model analyzing temporal signatures of allele frequency change between 1870-1960 to one that fits either two or three logistic regression lines between that time frame (i.e. a segmented logistic regression). A segmented logistic regression with two breakpoints provides the best fit to our data, compared to a model with either one or no breakpoints (two-break segmented AIC=54360.55, one-break segmented AIC =54437.66, non-segmented AIC=54444.67), and converges on 1913 and 1961 breakpoints, the later supporting a priori hypotheses and our interest in interrogating signals before and after the start of the green revolution in 1960 (**Fig 3D).**

We designed a randomization test based on observed allele frequency changes across the genome to obtain an expected distribution under null processes (drift, migration, and selection). In particular, we were interested in quantifying the potential bias in higher frequency agricultural alleles having the leeway to change more through time, as compared to a set of lower frequency alleles. We thus randomly sampled 154 loci from our contemporary collections (the same number as our observed clumped and historically matched set of agricultural alleles), 1000x across the genome, exactly matching the frequency distribution observed for extant agricultural alleles, first in extant agricultural and then in extant natural environments. To emphasize, this randomization was done independently in each environment, such that the alleles sampled to match the extant agricultural-allele frequency distribution in agricultural environments in one iteration were different from the alleles sampled to match the frequency distribution in natural environments (**Fig S4**). To account for the ascertainment bias in our set of putatively agriculturally adaptive alleles—finding alleles that show the greatest excess of allele frequency in agricultural compared to natural environments—we further constrained these randomizations to alleles across the genome which were at greater frequency in agricultural than in natural environments. On each of the 1000 randomizations within each environment, we then performed the same analyses as above: matching these alleles in our historical samples, producing a matrix of genotype data for both contemporary and historical sets, and performing a logistic regression for each locus, as well as logistic regression on all loci at once, for either samples from natural or agricultural environments, and for those that either preceded or were subsequent to 1960. For the 1000x randomizations within agricultural and natural environments, we then computed the 2.75 and 97.25% quantiles ("null 95% interquantile range") of the statistics of interest (total allele frequency change and selection coefficients) to compare against our observed

values.

Maximum likelihood estimate of selection

For the 7 known herbicide resistance alleles, we were particularly interested in individual estimates of selection on each allele over time. We used a maximum likelihood approach to estimate the strength of selection for each resistance allele between 1960-2018, along with a 95% credibility interval using profile likelihood. Summing over all years (*t*), the log-likelihood of observing the data is given by the binomial sampling formula describing the chance of observing the number of resistance (nR) and susceptible alleles (nS) in any given year:

$$ln(L) = \sum_t (n_R \cdot log \frac{p^{st}}{(1-p) + p^{st}}) + (n_S \cdot log \frac{p^{st}}{(1-p) + p^{st}})$$

where *p* represents the initial frequency of the allele when *t* = 0 (defined as the present) and *s* represents the strength of selection, both of which are unknown and estimated by maximizing the likelihood. Because many of the resistant alleles were only observed in contemporary samples, selection must be sufficiently strong to explain this rise, but the maximum strength of selection cannot be determined (the likelihood surface becomes flat). We thus only present the 95% confidence interval in the text (i.e., those values of the s for which the ln(L) falls within 12[0.05]/2 of the maximum likelihood). We implemented this algorithm in R, using the mle2 function implemented within the bblme package in R.

Ancestry inference

For genome-wide ancestry inference, we merged filtered SNPs from herbarium samples with high-quality SNP sets from (*11*) (n=187, collections from 2018) and (*31*) (n=162, collections from 2015), resulting in 457 individuals and representing all resequenced *A. tuberculatus* whole genomes. We used faststructure (*39*) to infer individual-level ancestry, taking the proportion of an individual's assignment to a grouping at K=2 to represent either var. *rudis* or var. *tuberculatus* ancestry. An individual's proportion of var. *rudis* ancestry was then analyzed in a multivariate regression that tested how well var. *rudis* ancestry was explained by longitude, latitude, environment (natural or agricultural), timespan (1800-1920 [n=39], 1920-1980 [n=44], 1920-2020 [n=374]), a two-way timespan by longitude interaction, a two-way timespan by state interaction, and a three-way timespan by environment by longitude interaction:

```
Individual ancestry assignment ~  longitude + latitude + environment +
timespan + timespan:longitude + timespan:state +
timespan:environment:longitude
```

We also used plink to perform a principle component analysis of merged SNPs from just herbarium samples (**Fig S12)** and all 457 samples jointly (**Fig S13).**

We were interested in the distribution of var. *rudis* ancestry across the genome, and so used LAMP (*40*) to assign ancestry to SNPs, based on two reference populations homogenous for either var. *rudis* or var. *tuberculatus* ancestry (Kansas and Ontario Natural Populations, respectively; (*31*)). Ancestry informative SNPs were those with an Fst > 0.40 (2x the mean genome-wide ancestry differentiation between varieties, in these two populations) between these reference populations and that were also in common between datasets (<20% of samples with missing data) after merging historical sequences with the contemporary paired sequence data (*11*). Since LAMP requires recombination rate information, we also imputed the LD-based genetic map from (*31*) to the ancestry-informative SNPs to get genetic distance between each. Finally, we performed the LAMP analysis, one population at a time, one scaffold at a time. After merging SNP-wise ancestry assignments across scaffolds, we calculated the mean, 5%, and 95% quantile of var. *rudis* ancestry in 100 kb regions for each population, and eventually, each environment (**Fig S10)**.

To understand the relationship between ancestry, agricultural selection, and genomic architecture, we performed a multiple regression to quantify drivers of fine-scale ancestry across the genome. We regressed the individual proportion of var. *rudis* ancestry in 100 kb windows across the genome against gene density, recombination rate, scaffold, environment, average CMH score, average XPEHH (difference in extended haplotype homozygosity across environments), the interaction between environment and average CMH score in each window, and the interaction between environment and the mean XPEHH in each window.

```
100kb mean ancestry ~  scaffold + mean_genedensity + mean_recomb +
mean_xpehh:env + mean_cmh:env + env
```

The least squares effect of environment on ancestry was taken to calculate the average difference in ancestry between agricultural and natural environments.

Tables S1 to S5

Table S1. GO Enrichment results for the top 0.01% CMH outliers (n=2055 SNPs).

GO biological process complete	Expected alleles	Fold Enrichment	Bonferroni p-value
growth	6.8	3.1	2.59E-02
anatomical structure morphogenesis	14.1	2.3	4.69E-02
anatomical structure development	44.1	1.9	1.67E-04
developmental process	46.2	1.8	3.20E-04
response to hormone	19.2	2.3	4.60E-03
response to organic substance	26.3	2.0	6.75E-03
response to chemical	41.4	1.8	1.79E-03
response to stimulus	85.3	1.6	3.30E-06
response to endogenous stimulus	19.6	2.2	3.38E-03
post-embryonic development	22.8	2.1	6.45E-03
multicellular organism development	38.2	1.9	7.02E-04
multicellular organismal process	42.5	1.8	6.59E-04
response to external stimulus	23.3	2.1	4.87E-03
system development	29.2	2.0	1.10E-03
response to abiotic stimulus	32.8	1.9	1.14E-02
response to stress	47.9	1.7	2.77E-02
cellular metabolic process	116.0	1.5	7.79E-07
metabolic process	127.9	1.5	5.08E-06
cellular process	182.1	1.5	5.85E-15
regulation of biological process	81.4	1.5	2.01E-02
biological regulation	89.9	1.5	1.05E-02
organic substance metabolic process	119.0	1.4	4.05E-04

Table S2. Gene and orthologue information for the 50 SNPs with the most significant CMH p-values, sorted by Scaffold and then CMH p-value. AMATA=*Amaranthus tuberculatus*, AT=*Arabidopsis thaliana*.

Scaffold	Position	CMH p-value	AMATA gene	AT gene	Orthologue	Blastn
1	59264411	6.08E-10	2592	NA	NA	Nuclear Fusion Defective 4-like
1	55158523	3.29E-09	2285	At4g34215	SGNH-hydrolase	
1	54510509	8.91E-09	2244	NA	NA	NA
1	12324718	1.93E-07	825	AT5G09550.1	Guanosine nucleotide diphosphate dissociation inhibitor (GDI)	
1	64789405	2.39E-07	3031	AT4G38380.4	Protein DETOXIFICATION 45%2C chloroplastic (DTX45)	
10	15265237	7.25E-08	22183	NA	NA	NA
10	26222541	1.03E-07	22540	AT3G29385.1	dentin sialophosphoprotein-like protein	
10	5374087	2.74E-07	21789	NA	NA	NA
11	24079642	8.55E-11	25990	AT5G63460.1	LOWER TEMPERATURE 1	
11	24078348	3.09E-10	25990	AT5G63460.1	LOWER TEMPERATURE 1	
11	24086055	1.10E-09	25990	AT5G63460.1	LOWER TEMPERATURE 1	
11	24006946	2.18E-09	25984	AT5G14220.4	**PPO2**	
11	24062979	2.36E-09	25989	AT5G50380.1	Exocyst complex component EXO70B1	
11	24080739	4.33E-09	25990	AT5G63460.1	LOWER TEMPERATURE 1	
11	32783081	7.71E-09	26623	NA	NA	
11	24048769	1.30E-08	25988	AT4G14110.1	COP9 signalosome complex subunit 8 (CSN8)	
11	24047019	4.13E-08	25988	AT4G14110.1	COP9 signalosome complex subunit 8 (CSN8)	
11	24083434	4.40E-08	25990	AT5G63460.1	LOWER TEMPERATURE 1	
11	24070607	4.79E-08	25989	AT5G50380.1	Exocyst complex component EXO70B1	
11	26024805	1.35E-07	26127	AT1G75125.1	plastid transcriptionally active protein	
11	25369807	1.47E-07	26088	AT5G39610.1	Nucleobase-ascorbate transporter 6 (NAC6)	
11	24021382	1.69E-07	25985	AT5G16550.1	LDAP INTERACTING PROTEIN	
11	24024155	1.69E-07	25985	AT5G16550.2	LDAP INTERACTING PROTEIN	
11	24048722	2.07E-07	25988	AT4G14110.1	CONSTITUTIVE PHOTOMORPHOGENIC 9	
11	24046969	2.55E-07	25988	AT4G14110.1	CONSTITUTIVE PHOTOMORPHOGENIC 9	
12	29335314	7.69E-10	24987	ATMG00310.1	ORF154	
12	29335422	3.09E-09	24987	ATMG00310.1	ORF154	
12	29327164	3.31E-09	24987	ATMG00310.1	ORF154	
12	29343299	1.81E-08	24987	ATMG00310.1	ORF154	
12	29336458	3.50E-08	24987	ATMG00310.1	ORF154	
12	29333763	5.83E-08	24987	ATMG00310.1	ORF154	
12	11427671	7.91E-08	24182	NA	NA	PPX2L
12	29328119	1.21E-07	24987	ATMG00310.1	ORF154	
12	11429925	2.10E-07	24182	NA	NA	PPX2L
12	29333575	2.30E-07	24987	ATMG00310.1	ORF154	
12	29333622	2.30E-07	24987	ATMG00310.2	ORF155	
13	35715956	8.71E-09	19321	NA	Calmodulin (Physarum polycephalum OX%253D5791)	
13	38847060	6.31E-08	19605	NA	NA	WIP2-like protein
13	26858220	8.73E-08	18867	NA	NA	NA
13	33084574	1.93E-07	19117	AT4G14310.2	KIN14B-interacting protein	
13	41600578	2.63E-07	19862	AT3G24160.1	PUTATIVE TYPE 1 MEMBRANE PROTEIN	
2	53114550	8.86E-08	5599	NA	NA	NA
2	10688517	9.38E-08	3997	AT3G09630.1	SUPPRESSOR OF ACAULIS 56 (SAC56)	
3	4797458	8.95E-09	6345	NA	NA	ATHB13
3	1072396	9.09E-09	5998	AT1G23820.1	SPERMIDINE SYNTHASE 1	
3	1072448	8.96E-08	5998	AT1G23820.1	SPERMIDINE SYNTHASE 1	
3	16438213	1.27E-07	7059	AT1G10150.1	Carbohydrate-binding protein	
3	5632229	1.80E-07	6414	AT4G09650.1	ATP SYNTHASE DELTA-SUBUNIT GENE (ATPD)	
6	18669007	1.54E-07	15047	AT4G35830.1	Aconitase 1 (ACO1)	
8	15398786	1.13E-08	20978	AT3G14310.1	Pectin Methylesterase 3	

Table S3. Selection-migration differentiation statistics for 6 common resistance alleles, along with estimates of selection estimated by logistic regression of the allele frequency through time. Ag, agricultural sites; Nat, natural sites. Cost and benefit estimates shown here for the additive (h=0.5) case. s (1960-2018) represents the maximum likelihood estimate of selection from the binomial sampling equation of allele frequency change, and we provide the associated 95% credible interval.

	Ag Freq	Nat Freq	Ag Benefit (s_A/m_A)	Nat Cost (s_N/m_N)	Cost/Ben	s (1960-2018)	95% CI of s
PPOdel	0.66	0.056	2.53119	10.7446	4.24488087	1.16	0.097, 3
EPSPSamp	0.48	0.25	2.16346	2.88	1.33120095	NA	NA
EPSPS106	0.89	0.08	0.40858	0.4884	1.19535954	1.12	0.053, 3
ALS122	0.034	0	NA	NA	NA	-0.01	-1, 3
ALS197	0.011	0.015	NA	NA	NA	0.286	-1, 3
ALS574	0.65	0.22	1.14286	1.51515	1.32575294	0.09	0.044, 0.0167
ALS376	0.921	0.0357	1.19023	2.51558	2.11352428	0.59	0.023, 3
ALS653	0.9326	0.0612	0.196636	0.215061	1.09370105	0.55	0.029, 3

Table S4. The top 15 loci with the strongest evidence of temporal selection between 1970 and 2018.

s	s p-value	s(SE)	AF change	scaf	position	AMATA annotated gene	AT orthologue	AT gene name
0.106	4.680E-04	0.030	0.846	Scaffold_11	26068182	AMATA_chromosomes_26131	NA	NA*
0.081	1.081E-04	0.021	0.823	Scaffold_2	10755264	AMATA_chromosomes_03999	Subtilase family protein	AT5G58840
0.057	2.382E-03	0.019	0.512	Scaffold_10	22995135	AMATA_chromosomes_22407	phosphotyrosyl phosphatase activator (PTPA family protein)	AT4G08960
0.052	8.211E-03	0.020	0.402	Scaffold_11	26024805	AMATA_chromosomes_26127	plastid transcriptionally active protein	AT1G75125*
0.052	6.084E-05	0.013	0.646	Scaffold_3	14213167	AMATA_chromosomes_06976	WRKY DNA-binding protein 13	AT4G39410
0.047	2.549E-07	0.009	0.783	Scaffold_10	36863790	AMATA_chromosomes_23250	hypothetical protein	AT1G36320
0.046	1.272E-01	0.030	0.162	Scaffold_3	49332690	AMATA_chromosomes_08117	NA	NA
0.045	2.889E-05	0.011	0.677	Scaffold_6	18669007	AMATA_chromosomes_15047	ACO1	AT4G35830
0.043	8.557E-05	0.011	0.599	Scaffold_12	21792253	AMATA_chromosomes_24760	NA	NA
0.043	4.601E-08	0.008	0.888	Scaffold_3	5632229	AMATA_chromosomes_06414	ATPD (F-type H+-transporting ATPase subunit delta)	AT4G09650
0.036	8.069E-06	0.008	0.666	Scaffold_12	5658420	AMATA_chromosomes_23853	NA	NA
0.033	4.729E-07	0.007	0.831	Scaffold_10	20690917	AMATA_chromosomes_22321	CCB2 (chaperone DUF2930)	AT5G52110
0.032	1.755E-07	0.006	0.819	Scaffold_10	16005835	AMATA_chromosomes_22202	NA	NA
0.032	3.939E-05	0.008	0.708	Scaffold_10	24312710	AMATA_chromosomes_22452	BPA4 (RNA-binding RRM/RBD/RNP motifs family protein, AT1G14340)	AT1G14340
0.031	5.175E-06	0.007	0.764	Scaffold_2	17891465	AMATA_chromosomes_04209	NA	NA

*** ~2 Mb from PPO**

Table S5. Metadata on herbarium collections.

Sample	Herbarium	Year	State	County/Locality	Description	Nat/Ag/Dist	Lat	Long	Catalog #
HBO900	INHS	1876	Illinois	Oquawka	Banks of the Mississippi	Nat	40.900098	-90.991298	1
HBO901	INHS	1870	Illinois	Kankakee	Wet Banks	Nat	41.154318	-87.919365	2
HBO902	INHS	1875	Illinois	Oquawka	Banks of the Mississippi	Nat	40.900098	-90.991298	3
HBO903	INHS	1897	Illinois	Warrenville	Mudflats	Nat	41.823199	-88.174393	6
HBO904	INHS	1892	Illinois	Chicago	Waste Ground ("Exposition Building")	Dist	41.779452	-87.61641	7 (429)
HBO907	INHS	1897	Illinois	Chicago	Dumping ground, Brighton Park	Dist	41.818787	-87.706608	10
HBO908	INHS	1952	Illinois	Fithian, Vermilion County	Along railroad	Dist	40.114645	-87.875056	11 (56279)
HBO909	INHS	1946	Illinois	Gardenplain, Whiteside County	Peat soil in potato field, "L.C. Anderson Farm"	Ag	41.782572	-90.138152	12 (19934)
HBO910	INHS	1948	Illinois	W of Antioch, Lake County	Disturbed soil	Dist	42.481444	-88.157622	13 (32479)
HBO911	INHS	1947	Illinois	W of Gillespie, Macoupin County	Corn field	Ag	39.127901	-89.860508	14 (22536)
HBO912	INHS	2005	Illinois	Minooka Rd & Route I-80, Minooka	Disturbed moist cropland margin near sable creek	Ag	41.457917	-88.308444	15 (242814)
HBO913	INHS	2000	Illinois	Cooperstown, Brown County	Former marsh in partly filled obow, mostly used as corn fields	Ag	39.959059	-90.611708	16 (205271)
HBO914	MBDH	1895	Ohio	Fulton Co.	Cornfield along the new river improvement on the prairie	Ag	41.67333333	-84.32694444	357
HBO915	MBDH	1903	Ontario	Essex Co.	NA	NA	41.772246	-82.79184	265
HBO916	MBDH	1955	Ohio	Greene Co.	Front lawn of Campus house	Dist	39.82083333	-84.01944444	292
HBO917	MBDH	1956	Ohio	Mercer Co.	Roadside, edge of soybean field	Ag	40.67861111	-84.51861111	197
HBO918	MBDH	1991	Ohio	Putnam Co.	Weedy ground alongside Blanchard river	Nat	41.03944444	-84.15694444	253
HBO919	IUH	1898	Indiana	Lake	on ballast (railroad bed)	Dist	41.602259	-87.25837	IND-0088703
HBO920	IUH	1941	Indiana	Vanderburgh	cultivated ground	Ag	37.870283	-87.634147	IND-0088729
HBO921	IUH	1941	Indiana	Vanderburgh	field north of river slough	Ag	37.979365	-87.544518	IND-0088730
HBO922	IUH	1952	Indiana	Johnson	Low cornfield	Ag	39.482726	-86.019624	IND-0088699
HBO923	IUH	2007	Indiana	Vigo	Mesic loam plain fallow cropland with successional weeds	Ag	39.43257	-87.380135	IND-0088738
HBO924	KUMH	1897	Missouri	Jackson, Courtney	Common along river	Nat	39.15556	-94.39333	176647
HBO925	KUMH	1913	Kansas	Doniphan,	Mesophytic oat field.	Ag	39.813187	-95.160615	43812
HBO926	KUMH	1913	Kansas	Doniphan,	Mesophytic oat field.	Ag	39.813187	-95.160615	43965
HBO927	KUMH	1949	Kansas	Shawnee, Lake Shawnee	Dry grassland.	Nat	39.000784	-95.626093	43811
HBO928	KUMH	1940	Kansas	Douglas, Lawrence	Field.	Ag	38.9365	-95.22391	44000
HBO929	KUMH	1970	Kansas	Chautauqua, Peru	Low, cultivated field. Gumbo soil.	Ag	37.09524	-96.07752	43956
HBO930	KUMH	1973	Kansas	Allen, Elsmore	Roadside right of way along field. Weedy, good soil.	Ag	37.79382	-95.16832	43987
HBO931	KUMH	1981	Kansas	Doniphan, Denton	Edge of cultivated field.	Ag	39.70289	-95.30739	43933
HBO932	KUMH	1994	Kansas	Labette, Parsons	Disturbed roadside along bean field and adjacent woodlands to the S. (scattered pop in field)	Ag	37.29416	-95.210223	111589
HBO933	ROM	1891	Ontario	NA	Leading to the H null Cemetery	Dist	43.892294	-81.312444	447
HBO934	ROM	1889	Ontario	Wingham	Alluvial soil (wet soil)	Nat	43.89161	-81.312795	189975
HBO935	ROM	1880	Ontario	London	Fields	Ag	42.954827	-81.234897	188334
HBO936	ROM	1891	Ontario	Leasselman	Alluvium along the Nation River	Nat	45.319337	-75.093578	438
HBO937	ROM	1880	Ontario	North Branch, London	Low River Flats	Nat	43.04267	-81.175264	441

HBO938	ROM	1940	Ontario	New Durham, Brant Co	Edge of Field	Ag	43.04956	-80.523938	129548
HBO939	ROM	1960	Ontario	Dunnville, Haldimand Co	Wasteground by Lake Erie	Dist	42.86879	-79.617923	159479
HBO940	ROM	1950	Ontario	Kent Co	Wet sand at Rankin creek near Mitchell Bay	Nat	42.48877	-82.413605	80.118
HBO941	ROM	1960	Ontario	Grenville Co, ON	Muddy shore of South Nation River	Nat	44.828922	-75.554629	134207
HBO942	ROM	1986	Ontario	Walpole Island	Disturbed forest/tall grass prairie - swampy	Nat	42.568369	-82.504545	TRT00029148
HBO943	ROM	1991	Ontario	Louth Twp., Niagara RM	W Shoreline of Jordan Marsh	Nat	43.174444	-79.375015	246184
HBO944	ROM	1985	Ontario	Ottawa-Carleton	Silty wet sand in dried-up pond in old pasture	Dist	45.274525	-76.088545	234833
HBO945	ROM	1988	Ontario	Wasaga Beach	Nottawasaga R. at Jack L.	Nat	44.485637	-79.99545	240977
HBO947	ROM	1997	Ontario	Brighton Tw., Northumberland Co	Sand pile, N shore of Lake Ontario	Nat	43.997787	-77.729332	256250
HBO948	ROM	1990	Ontario	Wingham	Riverbanks	Nat	43.89161	-81.312795	27813
HBO949	ROM	1880	Ontario	N.B. London	River flat	Nat	43.083152	-81.166767	85.246
HBO950	UMH	2001	Michigan	Lenawee Co	Open roadside at edge of cornfield, S side of Lime Creek Rd, w of Rogers Hwy	Ag	41.766065	-84.227971	1209687
HBO951	UMH	1909	Michigan	near Port Huron	Invading cultivated fields as a weed, abundant	Ag	42.979577	-82.470627	1209704
HBO952	UMH	1897	Illinois	Warrenville	Mud flats	Nat	41.829751	-88.177826	1559076
HBO953	UMH	1883	Ohio	Cincinnati	NA	NA	39.136053	-84.502294	1207957
HBO954	UMH	1833	Ohio	North Bend, Cincinnati	Sandy beach of Ohio River	Nat	39.14757	-84.753428	1207929
HBO955	UMH	1932	Ontario	Little Current, Manitoulin Island	In open fields	Ag	45.966667	-81.933333	1208039
HBO956	UMH	1986	Ontario	Essex Co, Malden Twp	Big creek marsh	Nat	42.05088	-83.056715	1208029
HBO957	UMH	1946	Illinois	Sangamon Co	Woods	Nat	39.578464	-89.730027	1208025
HBO958	UMH	1989	Ontario	N Plantagent Ip, Jessup's Falls (South Nation PP), Hwy 17 @ S Nation R	Rockshore protection and waste ground along river	Dist	45.559302	-75.064719	1207959
HBO959	UMH	1882	Illinois	Riverside	Gravelly, dry bed of Des Plaines River	Nat	41.826723	-87.82569	1559075
HBO961	UMH	1949	Illinois	Champaign Co	Mud and sand at bottom of a ditch	Dist	40.061911	-88.105443	1559082
HBO962	UMH	1993	Ontario	Middlesex Co, Lobo Twp,	Komoka Creek Swamp near Thames River, Edge of Thames	Nat	42.917996	-81.441417	1207956
HBO963	UMH	1995	Ontario	Niagara Regional Municipality, Niagara Falls Twp	Moist disturbed ground near railway tracks	Dist	43.06736	-79.084574	1208030
HBO964	MOB	1877	Illinois	Saint Clair	Low places	Nat	38.55	-89.916667	1740835
HBO965	MOB	1889	Missouri	St. Louis City	NA	NA	38.62975	-90.242434	1740859
HBO966	MOB	1893	Illinois	Saint Clair	Low places	Nat	38.509011	-90.177817	1740847
HBO967	MOB	1897	Illinois	Warrenville	Mudflats *note: same pop/date/collector as #93	Nat	41.829751	-88.177826	38973
HBO968	MOB	1902	Illinois	Saint Clair	Wet plains in East St Louis	Nat	38.625	-90.16	1740833
HBO969	MOB	1912	Missouri	St. Louis City	In open field	Ag	38.603097	-90.257021	778075
HBO960	MOB	1919	Illinois	Pope	Muddy banks of Ohio River	Nat	37.364829	-88.482798	853159
HBO970	MOB	1925	Illinois	Lake	Wet open ground, borders of tamarack swamp	Nat	42.370872	-88.121862	940825
HBO971	MOB	1927	Missouri	St. Louis	Mo River Sand Bar	Nat	38.595548	-90.767735	2157288
HBO972	MOB	1927	Illinois	St. Clair	In moist ground near Sugar Loaf Station, about 2 miles south of Du Po	Dist	38.495758	-90.217877	1026850
HBO973	MOB	1930	Missouri	St. Louis	Sandy fields	Ag	38.715451	-90.48898	2157285
HBO974	MOB	1933	Missouri	St. Louis	Alluvial banks along mississippi river, 6500 South St. Louis	Nat	38.572701	-90.229642	1140474
HBO975	MOB	1940	Illinois	Champaign	Moist thicket along sangamon River, 15 miles west of Urbana	Nat	40.119754	-88.498602	1579795

HBO976	MOB	1962	Illinois	Tazwell	Low River Bank	Nat	40.720818	-89.597212	1812554
HBO977	MOB	1967	Ohio	Lucas	Bottom of Maumee River at SW corner of Maumee city Limits	Nat	41.542549	-83.689433	6773122
HBO978	MOB	2005	Illinois	Madison	Weedy mowed and cultivated fields and bulldozed areas	Ag	38.779167	-89.640556	6113965
HBO979	MOB	2009	Missouri	St. Charles	Howell Island Conservation area, huge population, in large field next to hiking trail	Nat	38.665567	-90.70705	6449615
HBO980	MOB	2009	Illinois	Hancock	Dallas city public fishing access on the bank of the Mississippi River, and nearby railway tracks	Nat	40.634567	-91.179467	6443667
HBO981	MOB	2010	Missouri	Montgomery	Loutre River	Nat	38.911389	-91.592778	6334467
HBO982	MOB	2011	Illinois	Randolph	Roadside through floodplain with deep alluvial sandy loam	Dist	37.923333	-89.893333	6341594
HBO983	MOB	1919	Missouri	Jasper	A common weed in low and cultivated ground	Ag	37.204167	-94.344167	6773126
HBO984	AAFC	1892	Ontario	Ottawa	NA	NA	45.289379	-75.737232	1001151662
HBO985	AAFC	1955	Ontario	Richmond, Carleton	3 miles East, Wet Roadside	Dist	45.19477	-75.78076	1151663
HBO986	AAFC	1954	Ontario	Rockcliffe Park	Shoreline of Ottawa River, moist sandy gravelly soil	Nat	45.457179	-75.676253	1151664
HBO987	AAFC	1921	Ontario	Ottawa	Lac Constane	Nat	47.387012	-77.283366	1151665
HBO988	AAFC	1951	Ontario	Haldimand County	1 mile N of York, Sand beach	Nat	43.035303	-79.897787	1151667
HBO989	AAFC	1988	Ontario	Essex County	East beach, moist sand beach	Nat	41.93464	-82.505179	1151668
HBO990	AAFC	1998	Ontario	Essex County	North shore or W End, Moist woods	Nat	41.683999	-82.682764	1151669
HBO991	AAFC	1964	Ontario	Glengarry County	Dried bed of the Raisin River at rapids	Nat	45.13414	-74.575361	1151671
HBO992	AAFC	1949	Ontario	Lanarck	Along sandy beach, Between Mississippi River & McEwan Nay.	Nat	45.046655	-76.230997	1151672
HBO993	AAFC	1960	Ontario	Sarnia	High grass along tracks, Railyards north of the station	Dist	42.957915	-82.393943	1151673
HBO994	AAFC	1937	Ontario	Ottawa	Shore line, Moira Lake	Nat	44.486839	-77.458612	1151674
HBO995	AAFC	1960	Ontario	Grenville County	Muddy inundated shore of South Nation River	Nat	44.951292	-75.483802	1151675
HBO997	AAFC	1988	Ontario	Hald-Norfolk County	Rocky Point Prov. Park, Dunville Twp.	Nat	42.847407	-79.555433	1154001
HBO998	AAFC	1958	Ontario	Stormont County	Railway tracks	Dist	44.988422	-74.996022	1151676
HBO999	AAFC	1938	Ontario	Pelee Island	NA	Nat	41.683999	-82.682764	1154003
HB1000	AAFC	1934	Ontario	Mitchell's Bay	Waste places	Dist	42.468738	-82.407807	1154004
HB1001	AAFC	1939	Ontario	Waterloo County	Edge of 1st Island below Gault	Nat	43.352835	-80.316045	1154005
HB1002	AAFC	1918	Ontario	NA	NA	NA	42.77177	-81.197887	1154008
HB1003	AAFC	1961	Ontario	Waterloo County	Wet Field beside Speed River	Ag	43.455939	-80.290772	1154009
HB1004	AAFC	1828	Ontario	Middlesex County	Thames River	Nat	42.959472	-81.309866	1154010
HB1005	AAFC	1828	Ontario	Middlesex County	Bank of the Thames	Nat	42.980389	-81.344198	1154011
HB1006	AAFC	1961	Ontario	Waterloo County	Wet Field beside Speed River	Ag	43.455939	-80.290772	1154012
HB1007	AAFC	1966	Ontario	Middlesex County	Muddy soil along side creek	Nat	43.030996	-81.349254	1154014
HB1008	AAFC	1828	Ontario	Middlesex County	NA	NA	42.950068	-81.435938	1154013
HB1009	AAFC	1903	Ontario	NA	In boggy soil	Nat	43.892294	-81.312444	1151660
HB1010	INHS	1906	Illinois	Riverdale	Along railroad	Dist	41.640523	-87.626446	8 (2632)
HB1011	INHS	1894	Ilinois	Wheaton	Roadside	Dist	41.864251	-88.103346	9 (267)

Figures S1 to S13

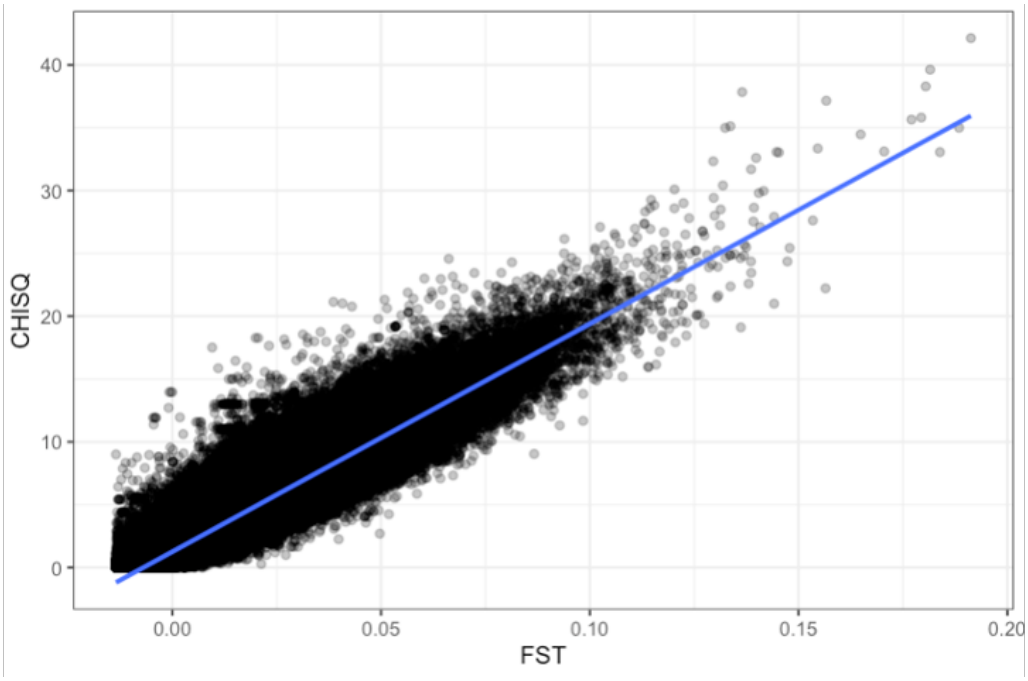

Fig S1. CMH X^2 statistic against between-environment F_{ST}, with the latter not stratifying for population pair.

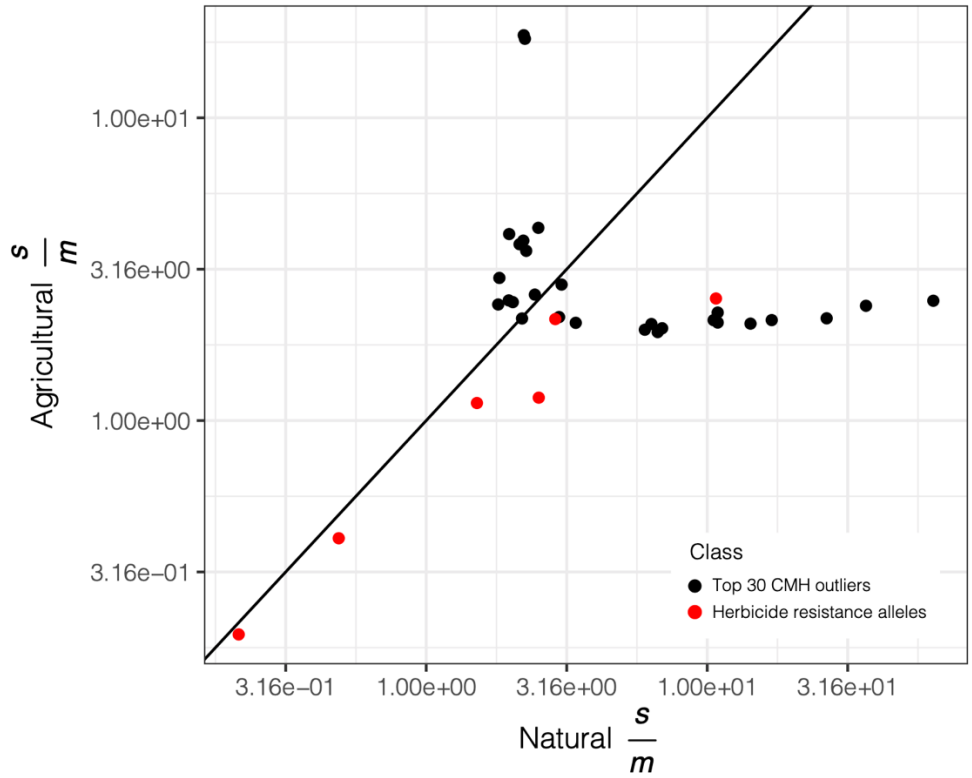

Fig S2. Agricultural $\frac{s}{m}$ versus natural $\frac{s}{m}$ for the 30 independent loci with the most signficant CMH scan hits compared to 6 common herbicide resistance alleles. Diagonal line represents equal agricultural benefits compared to natural costs, scaled by migration.

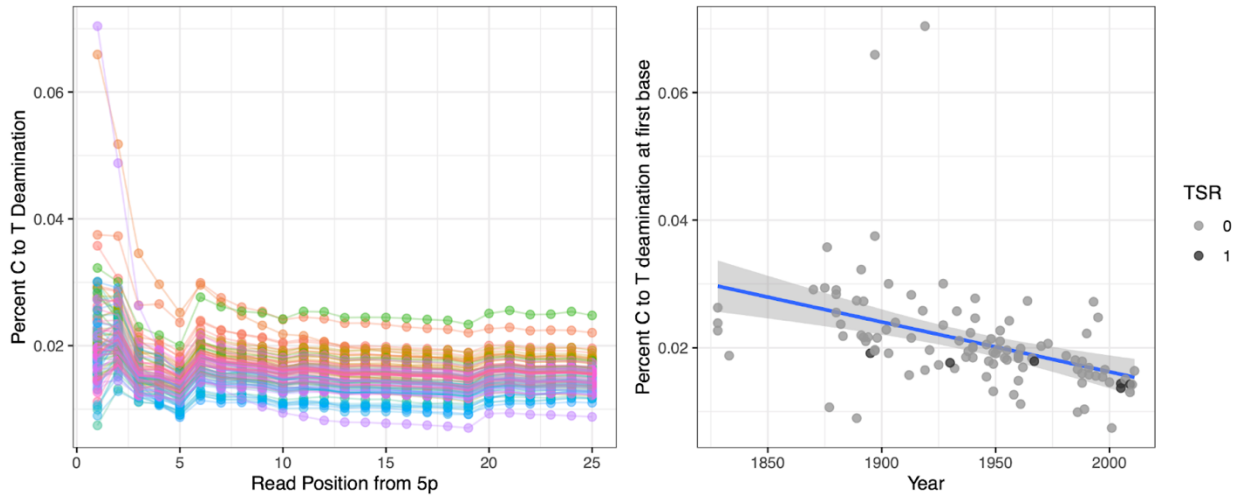

Fig S3. Percent C-to-T substitution at first base, and its correlation with read position and collection year for 108 sequenced herbarium samples. Right figure additionally illustrates C-to-T substitution level for the six samples with known target-site resistance alleles (three TSR genotypes excluded from selection analyses due to allelic bias).

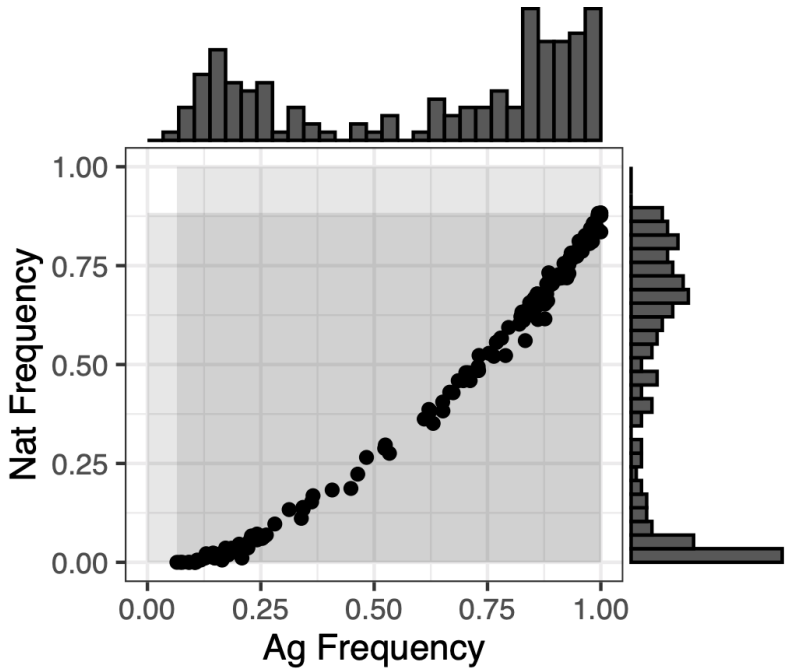

Fig S4. The distribution of frequencies for agriculturally adaptive alleles in agricultural samples along the x-axis, and in natural samples along the y-axis. Null distributions for an expectation of change in the frequency in our focal set of contemporary alleles was generated by producing randomized allele sets of the same size (n=154) matching the extant agricultural-allele frequency distributions shown here, first in natural environments (top histogram), and then in agricultural environments (right histogram).

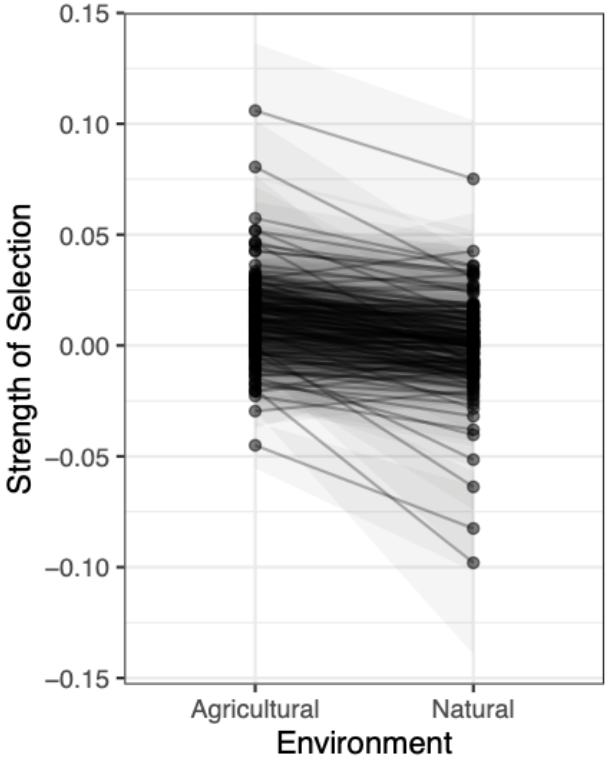

Fig S5. Inferred strength of selection on 154 agricultural alleles through time, in either agricultural or natural environments. Selection coefficients were extracted from logit-transformed logistic regressions of genotype on year, run separately for each locus in each environment. Gray ribbon for each locus represents the bounds of the standard error associated with the estimate of selection in each environment.

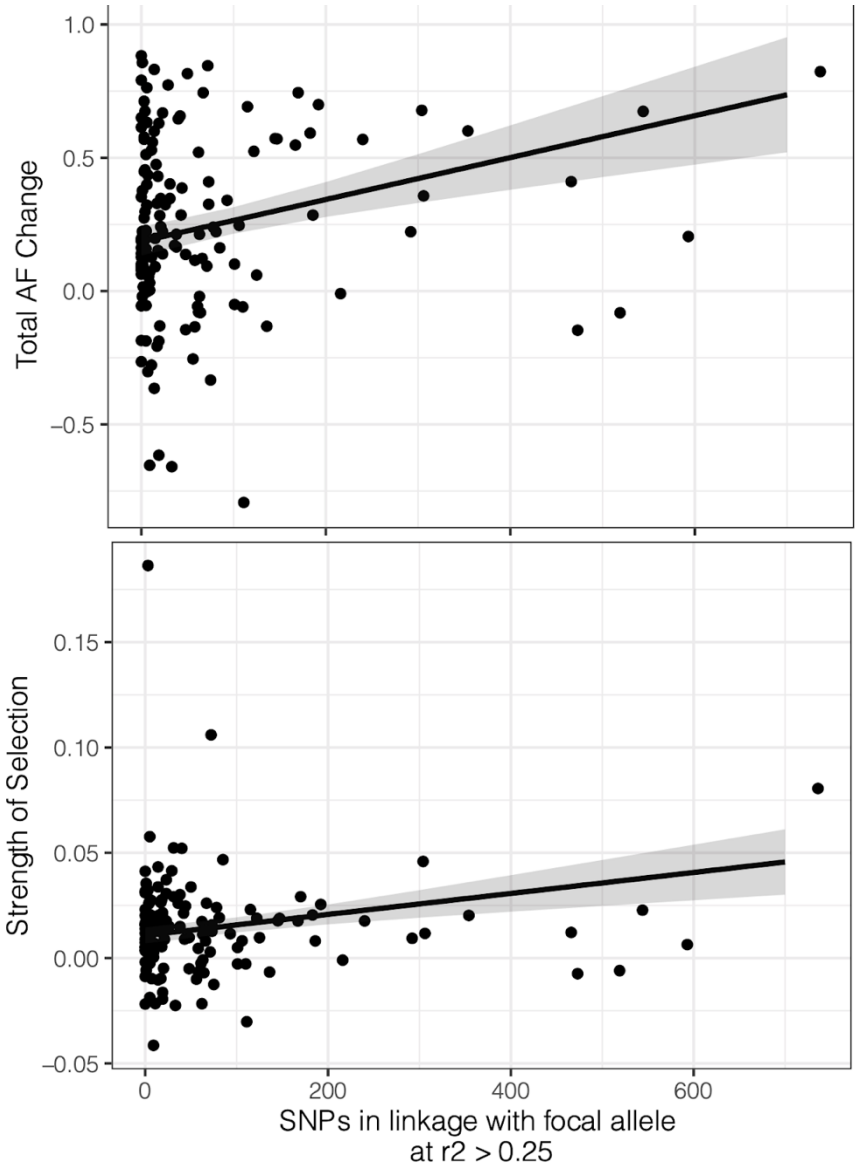

Fig S6. The association between contemporary patterns of linkage and selection and allele frequency change observed over the last 150 years across herbarium samples. Regression line shows the least square mean effect of contemporary linkage from a multiple regression analysis.

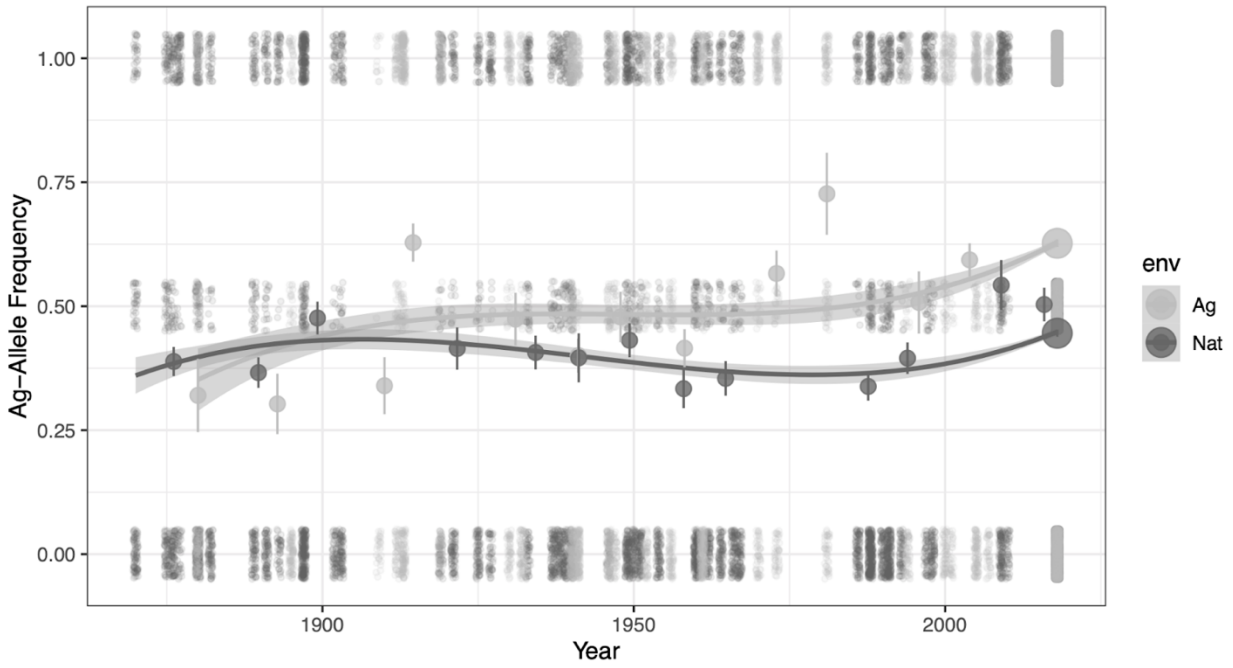

Fig S7. Cubic splines that illustrate the environment-specific frequency change of agricultural alleles through time since 1870. Gray ribbon denotes the 95% CI.

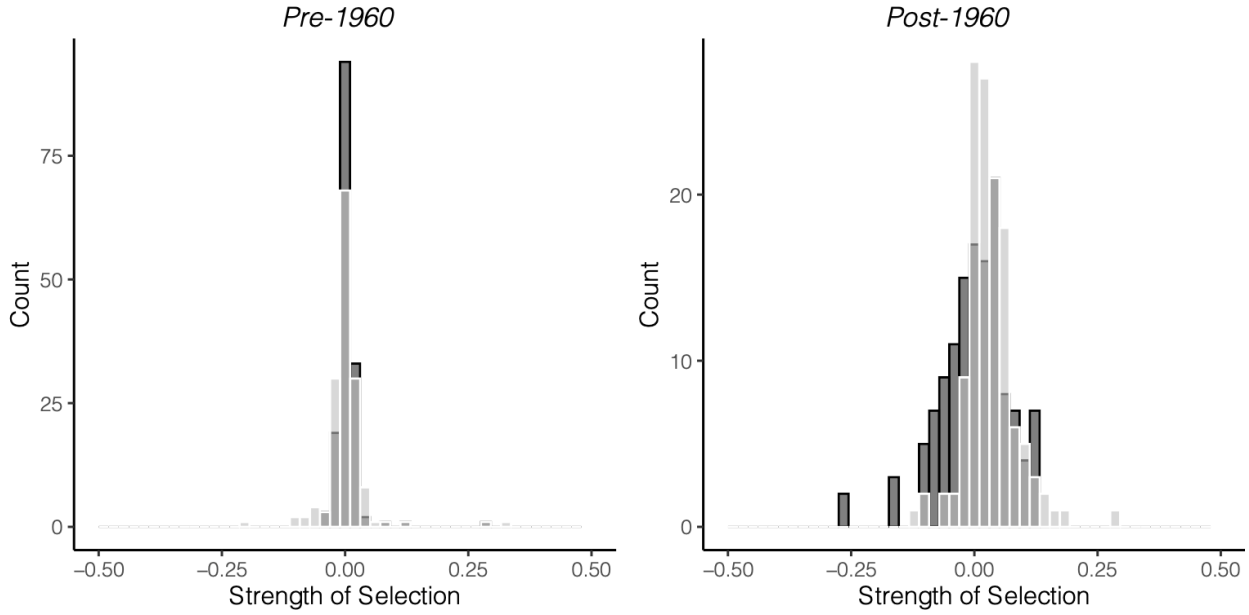

Fig S8. Logistic estimates of selection before (left) and after (right) the 1960s, the start of agricultural intensification, for agriculturally-associated alleles in natural (dark gray) versus agricultural and disturbed (light gray) environments.

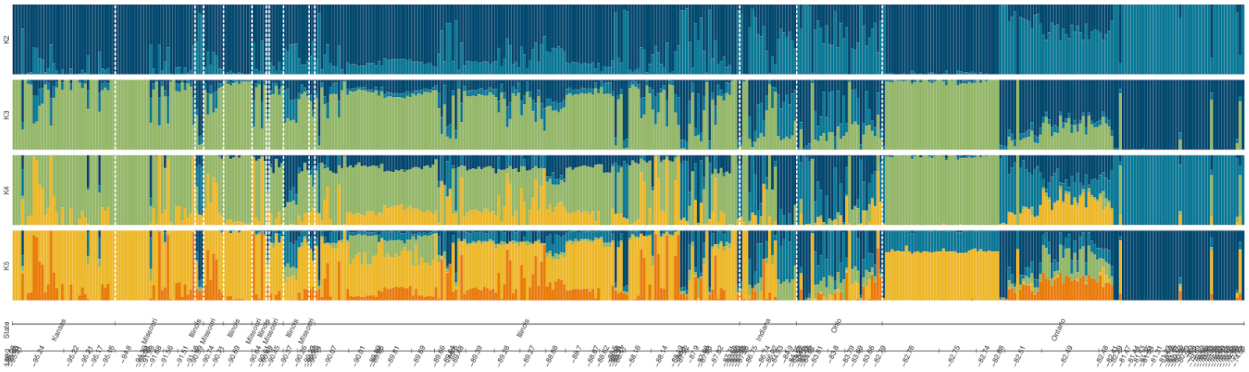

Fig S9. Longitudinal and state-wise patterns of ancestry across 457 *A. tuberculatus* individuals from contemporary and historical sampling, inferred from faststructure. Samples sorted by longitude, from west (left) to east (right). White dashed lines denote clusters of specimens sampled from different states and provinces across this longitudinal gradient. K=2 taken as var. *rudis* versus var. *tuberculatus* ancestry, as in (*31*).

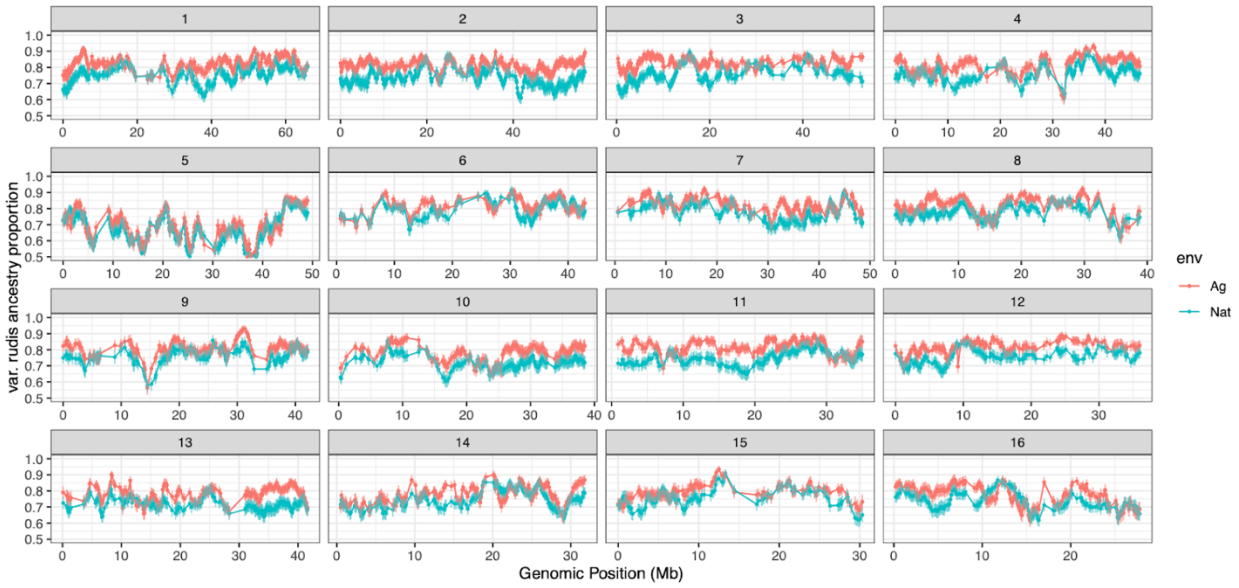

Fig S10. Excess of var. *rudis* ancestry in agricultural compared to natural environments, in 100 kb regions across the genome. Lines depict the mean ancestry across all populations within each environment, with error bars showing the mean 5th and 95th percentile of ancestry across populations. Fine-scale ancestry estimates were inferred with LAMP (*40*).

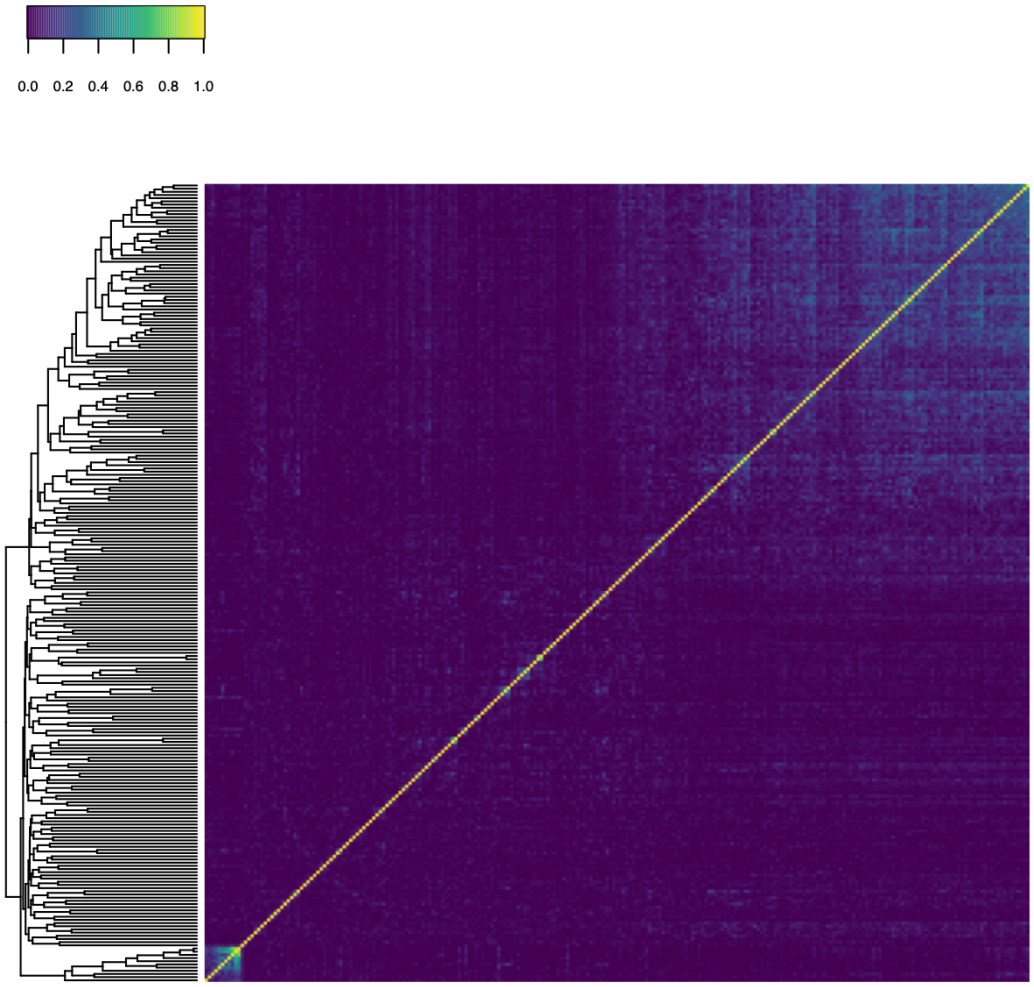

Fig S11. Heatmap of r^2 values alongside a dendrogram of the 254 agriculturally associated SNPs identified through CMH tests across paired contemporary natural-agricultural samples, illustrating independence among the 154 LD-clumped CMH outliers.

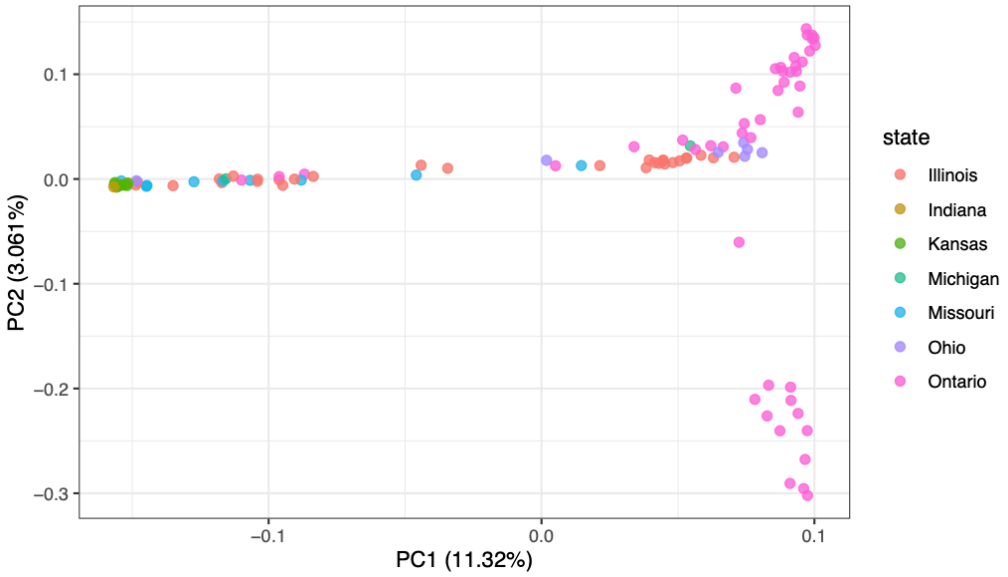

Fig S12. PCA of herbarium samples, coloured by state/province.

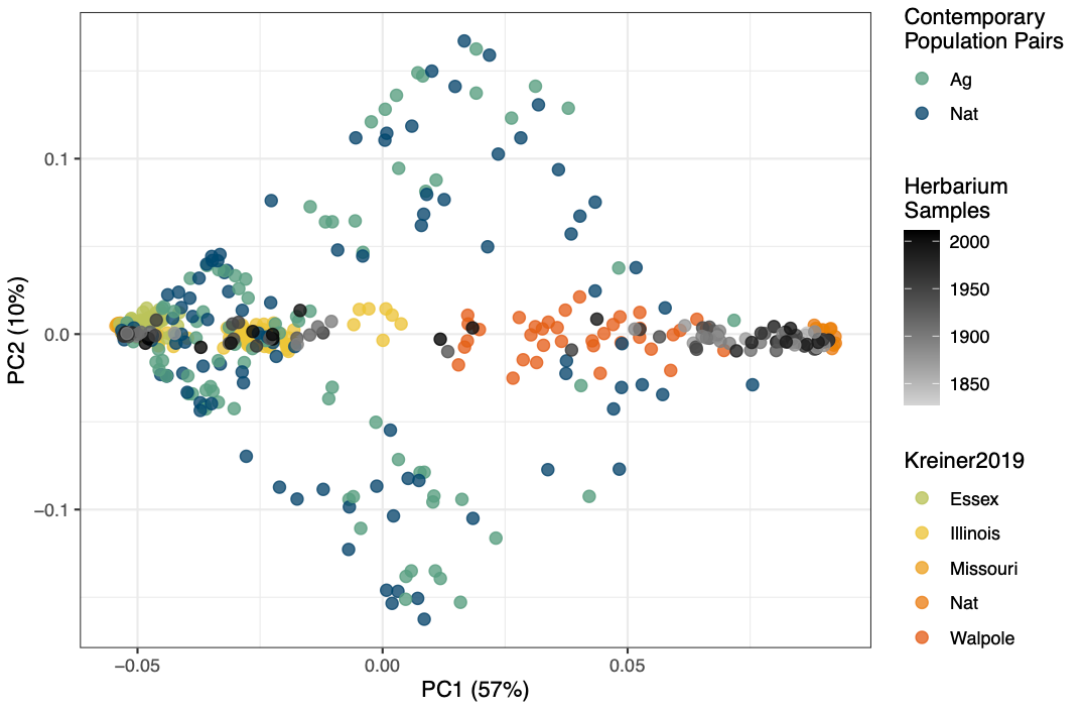

Fig S13. PCA of 457 *A. tuberculatus* specimens, including 108 herbarium samples along with contemporary paired populations (*11*) (n=187) and 21 populations from 5 geographic regions (*31*) (n=162).

Figure Captions

Fig 1. Sequenced waterhemp collections through space and time. A) Map of 14 contemporary paired natural-agricultural populations (n=187, collected and sequenced in Kreiner et al., 2021), along with 108 novel sequenced herbarium specimens dating back to 1828 collected across three environment types (Ag=Agricultural, Nat=Natural, Dist=Disturbed). **B)** Distribution of sequenced herbarium samples through time.

Fig 2. Signals of contemporary agricultural adaptation, gene flow, and antagonistic selection across the genome in *A. tuberculatus*. A) Results from Cochran–Mantel–Haenszel (CMH) tests for SNPs with consistent differentiation among environments across contemporary natural-agricultural population pairs. A 10% FDR threshold is indicated by the lower dashed horizontal black line, while the Bonferroni q-value < 0.1 cutoff is shown by the upper dashed horizontal gray line. Red points indicate focal adaptive SNPs after aggregating linked variation ($r^2 > 0.25$ within 1 Mb). Candidate agriculturally adaptive genes for peaks that are significant at a 10% FDR threshold shown. **B)** CHM results from the scaffold containing the most signficant CMH p-value, corresponding to variants linked to the PPO210 deletion conferring herbicide resistance and to the nearby herbicide-targeted gene *ALS*. **C)** Distribution of F_{ST} values between all agricultural and natural samples for ~3 million genome-wide SNPs (minor allele frequency > 0.05). Vertical lines indicate F_{ST} values for the 10 candidate genes named in A. **D)** Pairwise frequency of six common herbicide resistance alleles across agricultural and natural habitats sampled in 2018; the first four are nonsynonymous variants in *ALS* and *EPSPS*, the EPSPSamp is a 10 Mb-scale amplification that includes *EPSPS*, and the last one is an in-frame single-codon deletion in PPO (each dot represents on average ~5 individuals). Per migrant natural cost: agricultural benefit ratio relative to migration (C:B) is shown in the top right corner of each locus-specific comparison of frequencies across population pairs.

Fig 3. Genomic signatures of agricultural adaptation through time. A) Agricultural allele frequency trajectories for each locus, in agricultural and disturbed habitats (left), and natural habitats (right). Trajectories coloured by the quantile of frequency change in agricultural and disturbed habitats. Transparent lines indicate those with non-significant evidence of selection at α=0.05 after FDR=10% correction. **B)** The strength of selection on agricultural alleles for each locus in natural (dark gray) and agricultural and disturbed (light gray) habitats between 1870 and 2018. **C)** Agricultural allele frequency trajectories in each environment type, before and after the start of agricultural intensification in 1960. Vertical dashed line represents an inferred breakpoint in the data in a segmented regression. Environmental regression lines represent logistic fits to data that either predate or are subsequent to 1960. Large circles represent moving averages (over both loci and individuals) of allele frequencies, whereas dots represent raw genotype data for each locus and sample from which the allele frequency trajectory is estimated. Cropland use per capita in North America data from (*1*)**,** rescaled by use in 1600. **D)** The trajectory of alleles at known herbicide resistance loci through time, fit by logistic regression for each of the seven alleles present in our contemporary data. Dots represent genotypes for each historical and contemporary sample at each herbicide resistance locus. 95% credible interval of the maximum likelihood estimate of selection between 1960-2018 provided in the legend for each resistance allele.

Fig 4. Temporal shifts in the distribution of var. *rudis* ancestry have facilitated polygenic agricultural adaptation. **A)** Longitudinal clines in individual-level var. *rudis* ancestry over three timespans, illustrating the expansion of var. *rudis* ancestry eastwards over the last two centuries. **B)** The distribution of individual-level var. *rudis* ancestry by state and through time, illustrating state-specific changes in ancestry. Vertical lines represent first, second, and third quantiles of ancestry within each timespan and state. Timespans indicated in A) **C)** Increasing sorting of individual-level var. *rudis* ancestry into agricultural environments on contemporary timescales. **D)** Environment-specific metrics of selection (CMH p-value and cross-population extended haplotype homozygosity (XPEHH)) across the genome in 100 kb windows positively correlate with var. *rudis* ancestry in agricultural, but not natural habitats.