doi:10.1111/evo.14404

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

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

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

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

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

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

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

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

Received August 3, 2021
Accepted October 24, 2021

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

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

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

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

†These authors contributed equally to this work.

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

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

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

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

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

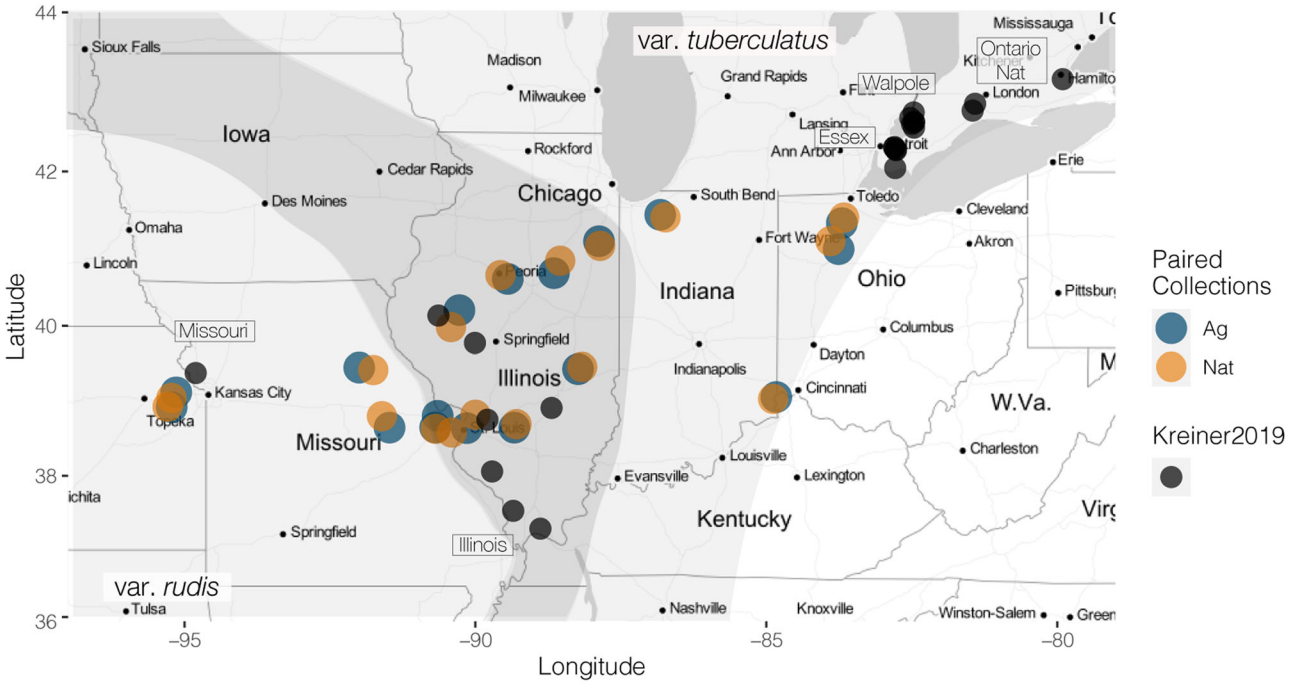

Figure 1. **Pairwise collections of natural and agricultural populations spanning the historical sympatric (dark shaded area) and allopatric ranges of *A. tuberculatus* var. *tuberculatus* (northeast) and *A. tuberculatus* var. *rudis* (southwest; range limits adapted from Sauer [1957]). For context,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

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

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

Methods

COLLECTIONS AND PARENTAL CROSSES

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

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

ROOFTOP COMMON GARDEN EXPERIMENT

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

PHENOTYPING AND DATA COLLECTION

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

DNA COLLECTIONS AND SEQUENCING

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

MAPPING AND SNP CALLING

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

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

POPULATION STRUCTURE

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

MIXED MODELS AND TESTS FOR PREADAPTATION

Modeling individual-level phenotypic variation

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

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

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

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

Testing the role of var. rudis *ancestry*

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

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

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

Results

DRIVERS OF THE DISTRIBUTION OF ANCESTRAL VARIATION

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

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

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

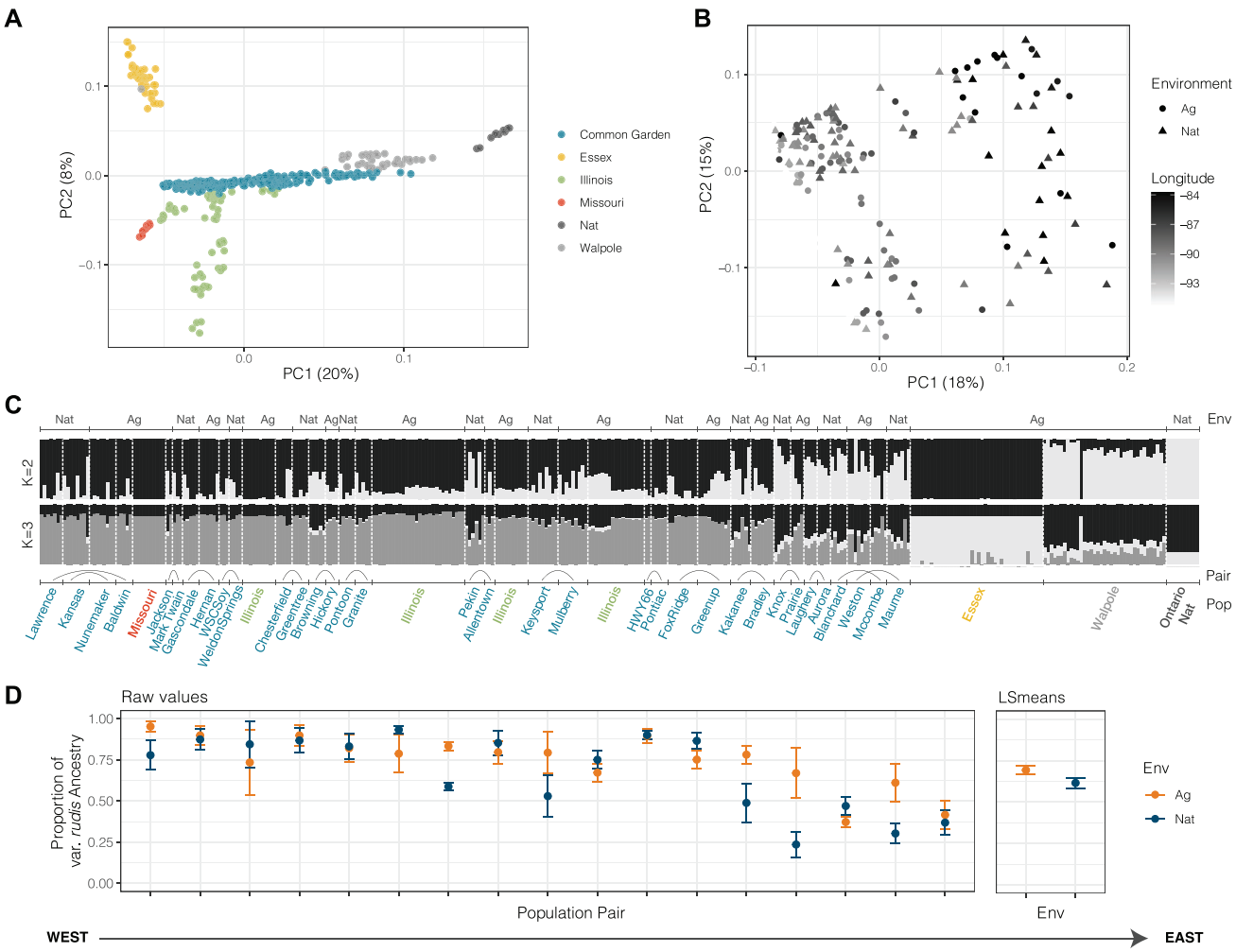

Figure 2. **Patterns of population structure from paired natural-agricultural collections, grown in the common garden experiment, in the context of samples from Kreiner et al. (2019). (A) Principal component analysis of samples from Kreiner et al. (2019), where populations have been identified as homogenous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

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

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

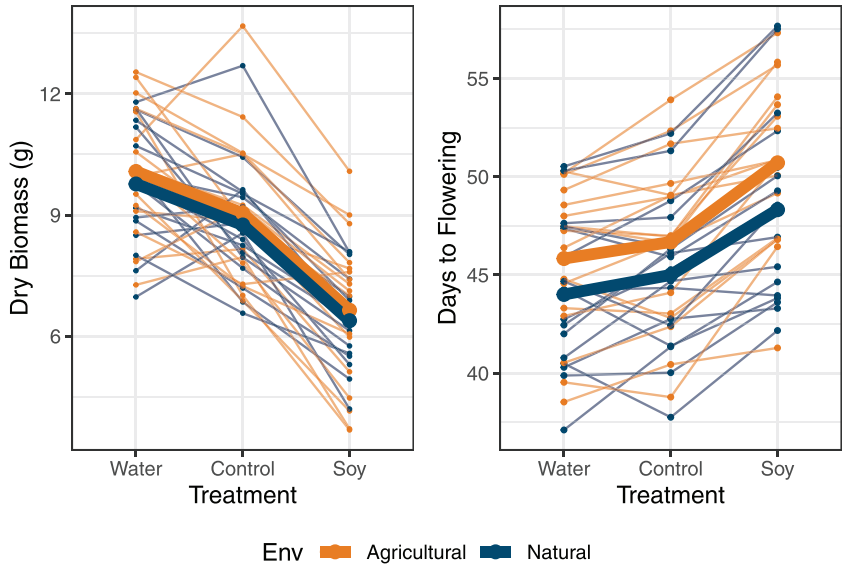

Figure 3. Population-level reaction norms of biomass and flowering time across the water, control, and soy treatments in the common garden experiment. Thin and thick lines represent population and environment mean reaction norms, respectively, and are additionally colored by whether collections were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

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

GENERAL OBSERVATIONS AND PLASTICITY IN A MANIPULATIVE COMMON GARDEN

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

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

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

DRIVERS OF PHENOTYPIC VARIATION AND THE ROLE OF ANCESTRY

Geographic, environmental, and ancestral trait divergence

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

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

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

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

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

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

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

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

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

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

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

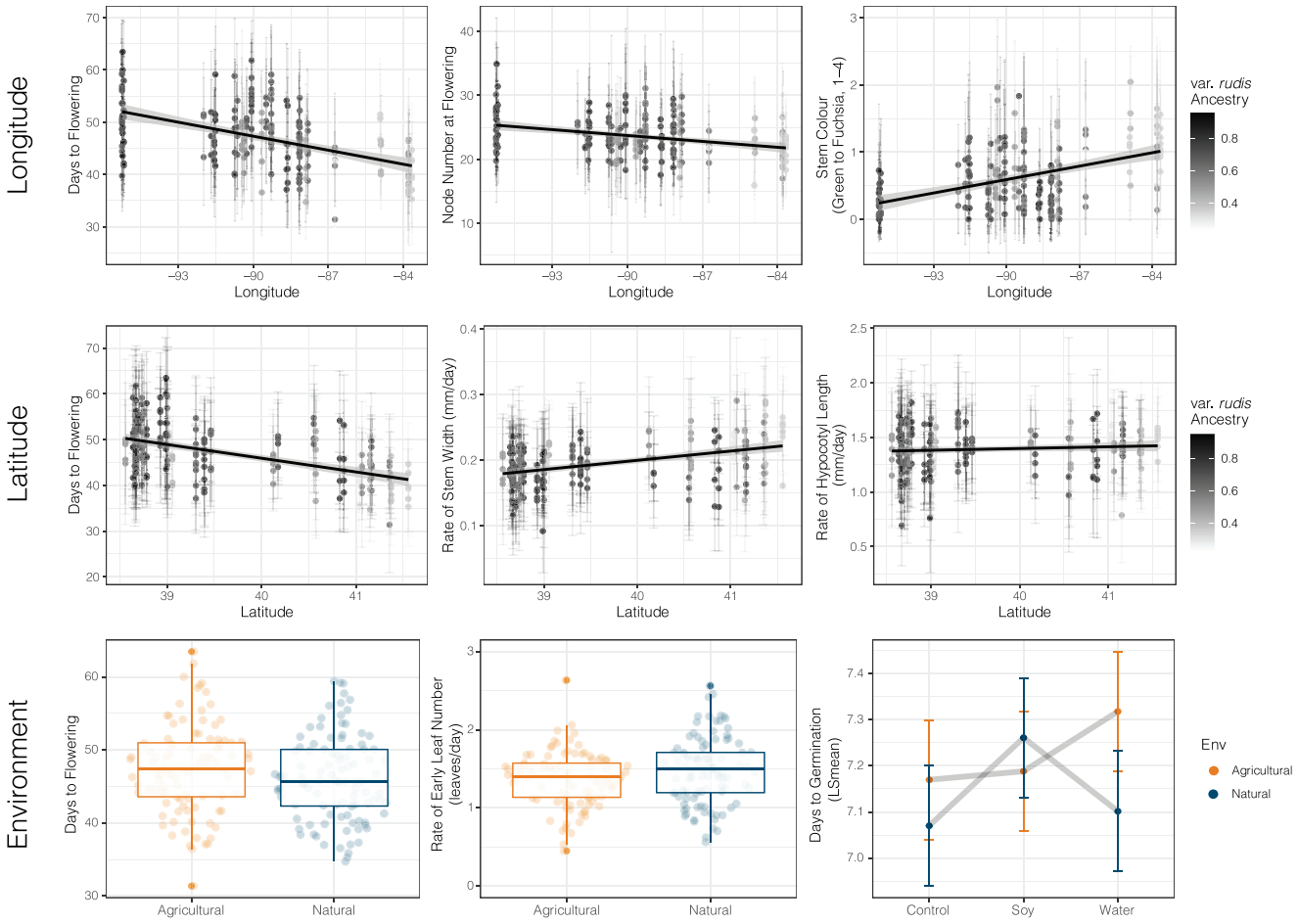

Figure 4. Phenotypic differentiation by longitude, latitude, environment, and the confounding effect of ancestry. All points indicate family-wise means, except for days to germination. (Top) Days to flowering, node number at flowering, and stem color by longitude and proportion of var. *rudis* ancestry.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

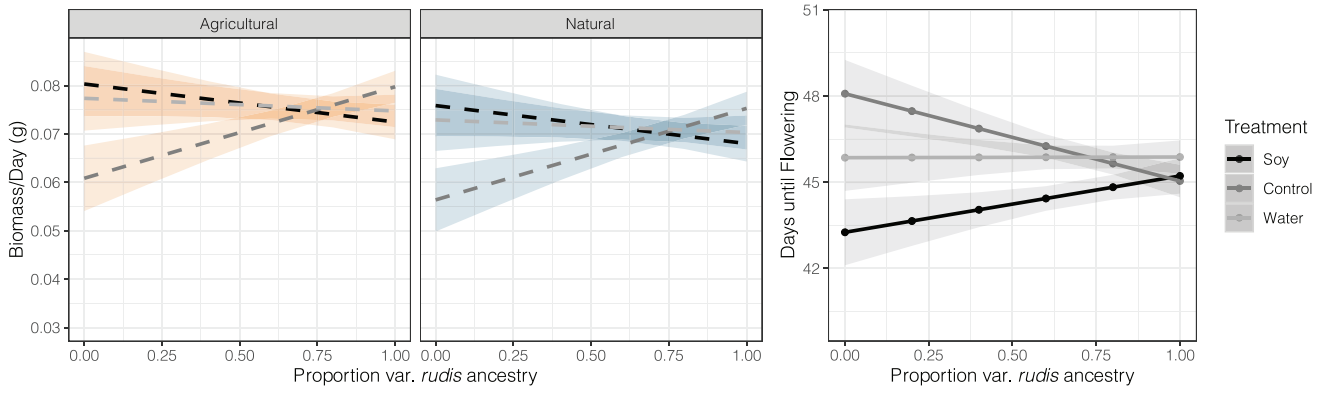

Figure 5. The treatment-dependent effects of var. *rudis* ancestry on male biomass (left panel) and male days to flowering (right). To the illustrate the significant linear, but not quadratic effects of ancestry, the illustrated values are based on the least-squares means of a reduced multiple regression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

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

Discussion

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

PHENOTYPIC UNDERPINNINGS OF AGRICULTURAL ADAPTATION IN THE FACE OF GENE FLOW

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

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

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

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

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

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

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

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

CONCLUSIONS

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

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGMENTS

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

DATA ARCHIVING

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

CONFLICT OF INTEREST

The authors declare no conflict of interest.

LITERATURE CITED

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

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

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

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

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

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

Broad-Institute. 2016. Picard tools.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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