Population genetics and origin of the native North American agricultural weed waterhemp (Amaranthus tuberculatus)

Katherine E. Waselkov and Kenneth M. Olsen

Published Version Information

Citation: Waselkov, K. E., & Olsen, K. M. (2014). Population genetics and origin of the native North American agricultural weed waterhemp (Amaranthus tuberculatus; Amaranthaceae). American Journal of Botany, 101(10), 1726-1736.

Digital Object Identifier (DOI): doi:10.3732/ajb.1400064

Publisher's Link:
http://www.amjbot.org/content/101/10/1726.full.pdf+html?sid=c3a78d54-2335-45ae-88e7-e7666bb3e117

Population Genetics and Origin of the Native North American Agricultural Weed Waterhemp (*Amaranthus tuberculatus*)[1]

Katherine E. Waselkov[2,3,4] and Kenneth M. Olsen[3]

[2]Current address: Division of Biology, Kansas State University, Manhattan, KS 66506 USA

[3]Department of Biology, Washington University, St. Louis, MO 63140 USA

[1]Manuscript received:___; revised manuscript accepted:___.

The authors thank P. Tranel at the University of Illinois Urbana-Champaign for *A. tuberculatus* transcriptome data; C. Lapid for help designing microsatellite primers; R. Kempen and N. Regenold for assistance with plant collecting; and L. Small, D. Vera-Vazquez, N. Regenold, A. Al-lozi, A. Clark, and D. Wetherington for assistance with lab work. Members of the Olsen, Miller, and Ferguson laboratories provided helpful comments on the manuscript. Funding was provided by an NSF Graduate Research Fellowship to K. Waselkov.

[4]Author for correspondence (email: waselkov@k-state.edu); present address: Division of Biology, Box 18, Ackert Hall, Kansas State University, Manhattan, KS 66506, USA.

ABSTRACT

Premise of the study—The evolution of invasiveness has been extensively studied in natural ecosystems; however, far less is known about the evolution of agricultural invasiveness, despite the major economic impact of weeds on crop productivity. Examining the population structure of recently arisen weeds can provide insights into evolutionary avenues to invasion of agroecosystems. Weeds that originate from wild plants are the most common yet least frequently studied type of agricultural invasive. Here we address several questions about the origin of the native North American agricultural weed waterhemp (*Amaranthus tuberculatus*), which invaded corn and soy fields in the Midwestern U.S. in the 20th century.

Methods—We genotyped 38 populations from across the species range with 10 microsatellite markers, and used these data to assess genetic diversity and population structure within and outside the geographical region where waterhemp is agriculturally problematic.

Key Results—We found evidence for two ancestral genetic lineages in our data, supporting the hypothesis that *A. tuberculatus* was diverging into two evolutionary lineages prior to the 20th century. However, we found no support for the hypothesis that agricultural weed populations arose from admixture of these two lineages after secondary contact. Our data suggest that eastward movement of the western genetic lineage, facilitated by changing agricultural practices, is the source of the agricultural invasion of waterhemp.

Conclusions—This research demonstrates that agricultural invasion by native, wild plant

species can proceed via different evolutionary trajectories from weeds related to domesticated plants, which has implications for evolutionary biology and weed control.

Key Words: agricultural weed; *Amaranthus tuberculatus*; hybridization; invasive; population admixture; waterhemp

93 The evolution of invasiveness, broadly defined as the ability of an organism to
94 establish, persist, and proliferate in a new habitat or range (Mack et al., 2000), has been
95 extensively studied in natural ecosystems (Lambrinos, 2004; Ward et al., 2008).
96 However, far less is known about the evolution of invasiveness in heavily human-
97 modified, anthropogenic habitats such as agricultural ecosystems. Agricultural weeds,
98 plants that invade crop fields and range lands, are the single greatest threat to agricultural
99 productivity worldwide, costing an estimated $33 billion per year in the US alone
100 (Pimentel et al., 2005). Examining the population structure and genetic composition of
101 these weeds can potentially provide insights into the evolutionary avenues to invasion in
102 agricultural habitats; this approach has been successful in natural ecosystem invasions
103 (Sakai et al., 2001). Despite the economic importance of agricultural weeds, population
104 genetics has been used to study the origin of very few weed species, aside from weedy
105 relatives of crops (Vigueira et al., 2013).
106 Agricultural weeds are often presumed to have evolved along with plant
107 domestication and the beginnings of agriculture roughly 12,000 years ago (De Wet and
108 Harlan, 1975). In such cases, the weed's origin in the distant past and long association
109 with highly mobile human populations can limit the ability to draw inferences about the
110 population-level processes underlying the evolution of weediness. However, agricultural
111 weeds may also evolve on a more contemporary timescale. The 20th century saw
112 enormous changes in agricultural practices in the U.S., including the introduction of
113 herbicides and the widespread adoption of conservation tillage (Owen, 2008), and these
114 changes have allowed species that were formerly confined to natural habitats to find a
115 new niche in an agricultural environment (Buhler et al., 1997). Unlike weeds whose

origins date to the beginnings of agriculture, recently arisen weeds may retain a clear genetic signature of the events that led to their agricultural invasion.

Three main hypotheses about the origin of agricultural weeds are prevalent in the literature (reviewed in Vigueira et al., 2013, following De Wet and Harlan, 1975). Weed species that are related to domesticated species may arise either through "de-domestication" (domesticated species becoming feral), or by hybridization between related domesticated and wild species. Support for these two hypotheses has been found in many systems, including beets, rye, rice, and sunflowers (Burger et al. 2006; Londo and Schaal, 2007; Olsen et al., 2007; Fénart et al., 2008; Muller et al., 2010). A close phylogenetic relationship between a crop species and a sympatric weed leads to interesting evolutionary dynamics, as ongoing gene flow between the two can shape adaptive evolution of the weed (possibly even through transgene escape), and many evolutionary studies have focused on these related crop-weed systems (eg., Warwick et al., 2003; Morrell et al., 2005; Aono et al., 2006; Campbell et al., 2006).

The third hypothesized mode of weed origination, the niche expansion of wild plants into agroecosystems through plasticity, adaptation, or exaptation, has received less attention by evolutionary biologists (but see Barrett et al., 1983; Menchari et al., 2007; Welsh and Mohamed, 2011), even though all weeds without close crop relatives must have followed this pathway to agricultural invasion, and even though this type of weed species is the most common (De Wet and Harlan, 1975). Of the few studies of this mode of agricultural invasion in the literature, the origin of weedy sunflower populations (*Helianthus annuus*) from wild populations is the best documented (Kane and Rieseberg, 2008; Lai et al., 2008). Due to the scarcity of studies on wild-to-weed evolutionary

transitions, many fundamental questions about these systems remain unaddressed, including: whether evolutionary changes in the plant and/or changes in agricultural management practices are required for the invasion; whether such invasions usually have a single or multiple wild sources; and what morphological, physiological, and ecological traits predispose a wild species to expand into agricultural habitats.

Among agricultural weeds that have evolved directly from wild species, a distinction can be drawn between those that are introduced in the region where they are invasive and those that are native. While agricultural weeds in North America are often assumed to be mostly introduced from elsewhere, approximately 50% of U.S. cropland weeds are in fact native (Clements et al., 2004). In many cases, wild populations of these native agricultural weeds still exist in the same geographic area, which opens up the potential for gene flow between populations in agricultural and natural habitats. This study focuses on one such native plant species that became agriculturally invasive over a very short, recent time scale.

Our study species, waterhemp (*Amaranthus tuberculatus* (Moq.) Sauer), is an herbaceous, outcrossing annual plant native to the Midwestern U.S., where it occurs naturally along riverbanks and in floodplains. Domesticated species of *Amaranthus* are largely absent from its range (Mosyakin and Robertson, 2003). Waterhemp has invaded Midwestern agricultural ecosystems since the 1950s and has become a major problem for corn and soybean farmers in Missouri, Iowa, and Illinois since the 1990s (Sauer, 1957; Tranel and Trucco, 2009). In Illinois alone, waterhemp accounts for about 10% of weed control costs for corn and soybean fields, costing farmers an additional $65 million per year (Patrick Tranel, Univ. of IL, pers. comm.). If uncontrolled, it can reduce corn yields

by up to 74%, and soybean yields by as much as 56% (Steckel, 2007). As a small-seeded annual with sporadic or discontinuous germination, waterhemp is a prime example of the class of agricultural weeds that benefited from the widespread adoption of conservation tillage in the late 20th century (Hager et al., 2000; Owen, 2008; Refsell and Hartzler, 2009). Other aspects of the species' biology have contributed to another factor in waterhemp's success: the rapid evolution and spread of herbicide resistance. A dioecious mating system and wind pollination are expected to promote extensive gene flow, potentially leading to genetic admixture and homogenization across large geographical areas. To date, resistance to five different chemical classes of herbicides has been detected in *A. tuberculatus* populations, and resistance to one of the older classes (ALS-inhibitors) is very widespread (Hausman et al., 2011; Tranel and Trucco, 2009).

Furthermore, there is some morphological evidence that the species may have been diverging into two species, one on either side of the Mississippi River, until human disturbance brought the taxa back into contact, and possibly gave rise to the agriculturally invasive strain through admixture (Sauer, 1957). Sauer studied herbarium specimens from 1856-1955, and noted that what he considered the "western species" had invaded the range of the "eastern species," with specimens appearing as far east as Indiana by the 1950s. He also found that hybrids between the two taxa were more likely to be collected in "artificial" habitats, including agricultural fields, than in natural habitats (Sauer, 1957).

In this study, we used population genetic techniques to address several questions about the origin and evolution of the agricultural weed form of this native Midwestern U.S. weed, to add to the small body of research on wild, native plants that have expanded their niche to encompass cropland. We extensively sampled populations of waterhemp

185 across the species' range, both within and outside of the region where waterhemp is
186 agriculturally problematic, and genotyped these populations using polymorphic
187 microsatellite markers. These population genetic data were used to test the following
188 hypotheses. First, we hypothesized (following Sauer, 1957) that *A. tuberculatus* was
189 diverging into two evolutionary lineages on opposite sides of the Mississippi River prior
190 to the 20th century, and that the present-day species would retain some genetic and
191 geographical signature of past subdivision into two evolutionary units.

192 The second hypothesis, contingent on the first, was that the agricultural weed
193 originated through hybridization between the two diverged lineages. Based on this latter
194 hypothesis, we predicted that populations of waterhemp collected from agricultural fields
195 would show strong evidence of admixture between western and eastern genetic
196 subpopulations. In order to test this prediction, we examined the genetics of agricultural
197 and non-agricultural waterhemp populations across the species' range and also in a
198 smaller geographical region at the range edge of weedy waterhemp in Ohio. Our data
199 support the first hypothesis but not the second, and suggest that eastward movement of
200 the western genetic lineage, facilitated by changing agricultural practices, is the source of
201 the agricultural invasion of waterhemp.

202

203 ## MATERIALS AND METHODS

204 ***Study system***—*Amaranthus tuberculatus* sensu lato (including *A. rudis* sensu
205 Sauer, 1972), is an annual herb native to North America. The species' range is centered
206 around the Mississippi Valley region, from the Great Plains (roughly as far west as the
207 100th meridian) eastward to Ohio, and from Louisiana northward to Minnesota, with a

northern range boundary in southern Ontario (Fig. 1). The region of agricultural invasion is more restricted: waterhemp is most problematic in the central Mississippi Valley region (MO, IL, IA, IN), but also occurs agriculturally in the eastern Great Plains and in parts of Kentucky and Ohio (Tranel and Trucco, 2009). Waterhemp is dioecious (and thus obligately outcrossing) and wind-pollinated, with small one-seeded utricle fruits that may be dehiscent or indehiscent. Natural populations of *A. tuberculatus* are almost always found in wet habitats, especially seasonally inundated riverbanks in the Midwest, but also banks of creeks and drainage ditches, lakeshores, and marshy floodplains (Mosyakin and Robertson, 2003).

Until Pratt and Clark's 2001 taxonomic study of populations across the species' range, waterhemp was considered two species, distinguished primarily by utricle dehiscence, tepal number, and geographic range: *A. tuberculatus*, the entity with indehiscent utricles and no pistillate tepals found to the east of the Mississippi River; and *A. rudis* (earlier misapplied name = *A. tamariscinus*; see Sauer, 1972), the dehiscent-fruited taxon with one pistillate tepal found natively west of the Mississippi River (Sauer, 1955; 1957; 1972; Fig. 1). Pratt and Clark discovered a continuum of morphological and isozyme characters across the range of both species and lumped them into the more broadly defined *A. tuberculatus*, but some authors still distinguish the two former species as varieties: *A. tuberculatus* var. *tuberculatus* and var. *rudis* (Costea and Tardif, 2003). The varietal taxonomy is used here, with Pratt and Clark's species called *A. tuberculatus* sensu lato (s.l.) or simply *A. tuberculatus*.

Sample collection—We collected plant samples from 38 populations of *A. tuberculatus* s.l. across the entire species range in 2009 and 2010 (Table 1 and Fig. 1;

Supporting information, Table S1). Populations were sampled in Nebraska, Kansas, Oklahoma, Arkansas, Missouri, Illinois, Indiana, Ohio, Michigan, and Ontario. Ohio was intensively surveyed because it is the edge of the range of agricultural waterhemp, with agricultural populations in ~10 counties west of Columbus but only non-crop populations in the remainder of the state (J. Stachler, NDSU, pers. comm.; Fig. S1). Gene flow between the weed and non-weed populations, which could obscure the genetic signature of the weed's origin, should be lowest in this area.

We located populations using a combination of herbarium record data and new surveys of typical *A. tuberculatus* habitat along riverbanks, lakeshores, and in crop fields. For the areas with agricultural waterhemp populations, both crop field and non-agricultural populations were included in the study. We recorded latitude and longitude coordinates for each population using a Garmin eTrex H handheld GPS unit (Garmin, Olathe, Kansas, USA), and collected voucher specimens (both male and female plants if possible). For each population, either 10 dried leaf samples in silica gel were collected, or 10 fresh leaf samples were collected and kept in a cooler until they could be frozen at -80ºC. Fruit dehiscence (considered an important taxonomic character for distinguishing the two varieties) was recorded for each female voucher specimen (Table 1).

DNA extraction and genotyping—DNA was extracted from each sample with Qiagen DNeasy Plant Mini Kits (Qiagen Inc., Valencia, California, USA). Ten microsatellite loci were amplified and genotyped. Primers, repeat motifs, and sizes of products are listed in Table S2. Multiple primer pairs from Lee et al. (2009), originally designed from *A. tuberculatus* genomic sequence data, were tested for consistent amplification and polymorphism, and three primer pairs were selected. The other seven

primer pairs were selected from a set of 14 mined from *A. tuberculatus* transcriptomic data using the program SSR Finder (Schroeder, 2003). The transcriptome contigs were provided by the Tranel lab (Univ. IL). In order to multiplex products from different primers in a cost-effective manner, we designed the forward primers with an M13(-21) sequence (TGTAAAACGACGGCCAGT) at the 5' end, to allow the attachment of a universal fluorescent-dye labeled M13(-21) tag (Schuelke, 2000). The universal tags were labeled with the fluorescent dyes HEX, 6FAM, and NED (Applied Biosystems, Carlsbad, California, USA). In addition, we designed the reverse primers to incorporate a PIG-tail, the sequence "GTTTCTT" at the 5' end of the reverse primer, to facilitate consistent non-template adenylation of the 3' end of the PCR product and to reduce stutter (Brownstein et al., 1996).

PCR was performed on ABI GeneAmp 9700 thermocyclers (Applied Biosystems), in 10 μL reactions containing: 1X GoTaq Flexi Buffer (Promega, Madison, Wisconsin, USA), 2.5 mM MgCl2, 0.05 mM each dNTPs, 0.15 μM M13(-21) dye-labeled tag, 0.04 μM forward primer, 0.16 μM reverse primer, 0.075 μL GoTaq (Promega), 3.875 μL nanowater, and 1.25 μL genomic DNA. Amplification conditions were: 94ºC for 5 minutes; then 30 cycles of 94ºC (30 seconds), 51ºC (45 seconds), 68 ºC (45 seconds) to amplify the product; followed by 8 cycles of 94ºC (30 seconds), 48ºC (45 seconds), 68ºC (45 seconds) to attach the labeled tag; and 72ºC (30 minutes) final extension. PCR products were diluted 1:10 and multiplexed (combining PCR products from up to three loci with different dye labels and different sizes in the same well) with 0.1 μL GeneScan 400HD ROX size standard (Applied Biosystems), denatured for 5

276 minutes at 95ºC, and genotyped on an ABI Prism 3130x Genetic Analyzer (Applied
277 Biosystems).

278 Microsatellite data were visualized using GeneMapper v3.7 software (Applied
279 Biosystems). The alleles at each locus for each individual were recorded by hand and
280 double-checked by repeated amplification and genotyping if more than two peaks
281 appeared (since *A. tuberculatus* is diploid) or unusual allele size classes were observed.
282 If these anomalies were observed twice (which happened very rarely for any particular
283 locus), the data for that marker for that particular individual were coded as missing.
284 Additionally, if genotyping failed for an individual for a particular locus, several
285 subsequent attempts were made to obtain these data before they were coded as missing.
286 Microsatellite data for all populations and loci is available from Dryad (XXX).

287 ***Microsatellite data analysis***—Microsatellite markers were checked for null
288 alleles using MICRO-CHECKER 2.2.3 (Van Oosterhout et al., 2004). The program
289 Genepop 4.2 (Rousset, 2008) was used to test the probability of Hardy-Weinberg
290 equilibrium for each population (with the Markov chain method to estimate exact p-
291 values), to test for linkage disequilibrium between loci (with Fisher's method), and to
292 detect private alleles for each population. Popgene 1.31 (Yeh et al. 1997) was used to
293 estimate the average number of observed alleles and effective alleles and average
294 observed and expected heterozygosity over all loci for each population, and per-locus for
295 all populations combined. Weir and Cockerhams's theta (an estimate of Fst) was
296 calculated over all loci and all populations using the program FSTAT 2.9.3.2 (Goudet,
297 1995).

298 To test for isolation by distance (IBD), we calculated the geographic great circle
299 distance in kilometers between each pair of populations using the Geographic Distance
300 Matrix Generator (Ersts, 2013). We then generated a matrix of pairwise Fst values
301 between populations, and compared the two matrices in a Mantel test for isolation by
302 distance (with 1000 permutations) using Genepop. This procedure was used to test for
303 IBD across the entire species range, and across subsets of the species range: the Plains
304 states (TR, CHE, TCL, and SaltR populations); Missouri and Illinois; Ohio; and Northern
305 Ohio (OTT, MAU, and PTC populations), Michigan, and Ontario.
306 To identify the highest-likelihood number of genetic subpopulations (K) in the
307 data, we employed the program STRUCTURE 2.3.1 (Pritchard et al., 2000). We used the
308 correlated allele frequencies model with sampling locations (population assignment) as a
309 prior, which facilitates clustering for data with weak genetic structure. We ran separate
310 analyses for the admixture and no-admixture ancestry models for both datasets. For the
311 total species range dataset, we ran the analysis for K=1-20; and for the Ohio dataset, from
312 K=1-7, with five runs per K, 100,000 Markov Chain Monte Carlo (MCMC) burnin steps,
313 and 500,000 MCMC steps after the burnin for both datasets. The separate analysis of the
314 Ohio dataset was conducted to examine fine-scale structure in this intensively sampled
315 part of the range. To estimate the number of genetic clusters from the ln Probability
316 (X|K) values output by STRUCTURE, we used STRUCTURE HARVESTER v0.6.93
317 (Earl and vonHoldt, 2012) to implement the ΔK method of Evanno et al. (2005). We
318 combined the results of multiple runs for each K, using the *FullSearch* algorithm, with
319 the program CLUMPP v1.1.2 (Jakobsson and Rosenberg, 2007) and visualized the
320 combined results as bar graphs with *distruct* v1.1 (Rosenberg, 2004).

The program BAPS v5.3 (Bayesian Analysis of Population Structure; Corander et al., 2003) was used to incorporate spatial information into the estimation of genetic clusters for the total species range dataset, using an admixture model rather than a no-admixture model, as recommended by François and Durand (2010). Unlike STRUCTURE, BAPS does not use MCMC to infer K. Instead, BAPS uses a stochastic search algorithm that considers multiple K values simultaneously to directly estimate the number of genetic clusters and assign individuals to those clusters using mixture analysis. Geographic localities of populations were employed as priors, using the "spatial clustering of groups" option (Corander et al., 2008). We set K_{max} = 2, 5, 10, and 20, with three runs per K. For admixture analysis based on the mixture analysis results from the highest-probability K-value, we used a minimum population size of five, 200 iterations, 100 reference individuals for each population, and 20 iterations of reference individuals (as suggested in the manual). The graphical output from BAPS is an "admixture partition" bar graph showing genetic assignment of individuals, and a Voroni tessellation diagram showing the spatial genetic assignment of populations.

RESULTS

Genetic diversity—Populations largely conformed to Hardy-Weinberg expectations. MICRO-CHECKER revealed that none of the loci were consistently more homozygous than expected, indicating no evidence for null alleles in the dataset. Likewise, no single population showed a deviation from Hardy-Weinberg equilibrium (HWE) at more than one locus. Hardy-Weinberg probability tests yielded slightly different results, showing that the locus *AAC1* deviated from HWE in five populations,

and that one population, MAU, deviated from HWE when all loci were taken into account (P = 0.024). Population structure analyses were therefore run with and without this locus and population; however, exclusion of these data had negligible effect on the results (data not shown). No significant linkage disequilibrium was detected between any pair of loci.

Genetic diversity data for each population are presented in Table 2, with per-locus values shown in Table S3. Gene diversity (He) ranged between 0.42 and 0.68 for individual populations, with a mean of 0.56 over all populations, indicating high within-population genetic diversity. Populations in the western half of the species range tended to have higher average observed and effective numbers of alleles than populations in the eastern half of the species range. Several populations in Indiana (VIGO, KANK) and Ohio (NEV, CAN, SCIO, MC, PTC) and one population in Illinois (PEK) had higher average observed than expected heterozygosity, potentially suggesting recent admixture (Table 2). These populations include only two of the ten populations from crop fields (see Table 1), and agricultural waterhemp populations did not have higher average observed than expected heterozygosity on average (results not shown). The number of alleles per locus ranged from 6 to 20, and in general both observed and expected heterozygosity per locus were high, with the exception of the locus *ATC9*, which had approximately one effective allele (Table S3).

Genetic differentiation among populations was low overall. Weir and Cockerham's theta (analogous to Fst) was 0.075 averaged across loci, with a range of 0.029-0.186 for individual loci. Nonetheless, Mantel tests revealed significant isolation by distance across the species range ($P < 0.00001$, Fig. 2), with pairwise Fst values

between populations ranging from 0.0013 to 0.2681. For geographical subsets of the species range, there was no isolation by distance at the state or bi-state level, or across the three Plains states. However, the dataset composed of populations from northern Ohio, Michigan, and Ontario did show weak isolation by distance (P = 0.037).

Population structure—

Whole species range dataset—STRUCTURE output and the ΔK statistic indicated two genetic clusters within *A. tuberculatus*, one characteristic of the western part of the geographic range and one characteristic of the eastern part, with greater than 25% admixture inferred for the populations PEK and KNK (IL), KANK, IND, and WAB (IN), SCF (ON), and the OH agricultural waterhemp region populations (Fig. 3; Supporting information, Fig. S2). Overall for K=2, the ten crop field populations had an average of 23.9% admixture, while the 28 non-agricultural populations had an average of 15.0% admixture (calculated from STRUCTURE output). Analyzing the dataset with the no-admixture model yielded the same inferred numbers of clusters (results not shown).

The analysis of population structure that explicitly incorporated information on the geographical locations of populations yielded generally similar results to the STRUCTURE analysis. BAPS identified the highest K value as K=3 (for K_{max} = 5, 10, and 20). Population and individual assignments to these three clusters are shown in the admixture partition bar graph (resulting from a mixture analysis, followed by admixture analysis) (Fig. 4a). The population assignments largely correspond to the same western/eastern divide seen in the STRUCTURE analysis, with the switch from primarily western to eastern genetics between VIGO (IN) and KANK (IN), PEK having a strong affinity for the eastern genetic cluster, and IND, RT29, GTB, and STW primarily

assigned to the western cluster. The MC population (OH) was the only population assigned to the third cluster. BAPS identified only eight individuals as exhibiting admixture, but these individuals were in the central populations (PEK, KNK, KANK, IND) and agricultural waterhemp region of Ohio (STW, SCIO), which correspond to populations with high admixture in STRUCTURE. The Voroni tessellation diagram shows the same population genetic assignments spatially (without admixture) (Fig. 4b). The agreement between the STRUCTURE and BAPS results suggests that a genetic signature of the two previously-diverging lineages in the species still remains in the present-day species.

Ohio dataset—For the dataset composed of only Ohio population genetic data, the ΔK method of Evanno et al. (2005) (implemented in STRUCTURE HARVESTER) weakly supported two genetic clusters, one in the agricultural waterhemp region of Ohio and the other in the Ohio River region, with strong admixture between the two clusters inferred for populations in northern Ohio and in the southern Ohio population BTL (Figs. 5, S3). In contrast to the whole range dataset results, three of the four agricultural populations in the Ohio dataset (GTB, MC, and RT29) have less than 25% admixture in the STRUCTURE bar graphs for K=2: instead, most individuals in these populations were assigned almost entirely to the western genetic cluster (Fig. 5). Overall for K=2, populations in the agricultural waterhemp region had an average of 12.6% admixture, while the Ohio populations outside of this region had an average of 31.8% admixture. There was a second peak in ΔK at K=5 for the Ohio dataset (Figs. S3, S4).

DISCUSSION

Support for two ancestral evolutionary units— The combined results from isolation by distance tests, STRUCTURE, and BAPS analyses are largely congruent and paint an interesting picture of the recent origin and evolution of the agricultural weed form of *Amaranthus tuberculatus*. The hypothesis that the species was formerly diverging into two evolutionary units prior to the 20th century is supported by our data. STRUCTURE recovered two genetic clusters from the total species range dataset, and at the range edges, the geographical structure of these clusters corresponds closely to the hypothesized eastern/western divide between the two former taxonomic units (Sauer, 1957) (Figs. 1, 3). BAPS identified an additional cluster consisting of one Ohio agricultural population, which might have distinctive multigene allele frequencies due to admixture. It appears from our data as though the Mississippi River is no longer the geographical divide between the two genetic clusters; instead, the western genetic cluster extends into Indiana. Interestingly, this boundary shift was documented more than half a century ago by Sauer (1957), who observed from herbarium specimen records that the western taxon, now called *A. tuberculatus* var. *rudis*, had been moving steadily northward and eastward across the Mississippi River since the 1850s, into the range of *A. tuberculatus* var. *tuberculatus*. Furthermore, he noted that this movement was associated with agricultural invasion: the earliest records of *A. tuberculatus var. rudis* in Illinois (1940s) and Indiana (1950s) are reports from agricultural fields (Sauer, 1957).

Spatial genetic clustering allows the use of information beyond genotype data (such as spatial autocorrelation and geographical trends) for inferring population structure, and can be especially useful when closely-related taxa come into secondary contact at regional geographic scales (François and Durand, 2010). The differences in

436 clustering between STRUCTURE and BAPS are probably due to differences in the
437 assumptions of the underlying Bayesian clustering methods, as well as the inclusion of
438 geographical information. While STRUCTURE uses Markov methods to find the highest
439 likelihood for each K value independently, BAPS uses a stochastic search algorithm that
440 directly estimates the most likely K value. Furthermore, BAPS estimates admixture after
441 partitioning the data into clusters with a mixture model, while STRUCTURE estimates
442 admixture and the likelihood of each K value simultaneously. Because of these
443 differences, the discovery by both programs of two major genetic clusters in our dataset
444 provides strong support for this inference.

445 Pratt and Clark's (2001) analysis of 27 morphological characters and 14 isozyme
446 loci across the range of *A. tuberculatus* s.l. revealed a continuum of morphological
447 character states and isozyme alleles across the entire range. On the basis of no clear
448 clustering in a PCA of these characters, they concluded that the two previously described
449 taxa are one variable species. The observed continuum is not surprising, given the
450 geographical overlap between the two varieties that occurred in the middle of the range as
451 *A. tuberculatus* var. *rudis* pushed eastward. In our own voucher specimens, the
452 morphological character of utricle dehiscence is nearly constant at the western and
453 eastern ends of the range (the Plains states and Ontario), and variable among populations
454 in the range center (Table 1). Both Sauer and Pratt and Clark were primarily focused on
455 taxonomy, and tended not to focus on population-level patterns. Our application of a
456 relatively recently-developed genetic tool, microsatellite genotyping, has largely
457 confirmed their broad-scale observations and has also shed more light on the origins of
458 the agricultural weed form.

Intraspecific hybridization did not create the agricultural form—Our second hypothesis, that weedy waterhemp was created through hybridization between the two evolutionary units in *A. tuberculatus* s.l., was not supported by our data. If this were the case, one would expect that the agricultural populations of waterhemp would show strong evidence of admixture between the two genetic clusters. Instead, four out of the six Missouri, Illinois, and Indiana agricultural populations show a strong affinity with the western (red) genetic cluster in STRUCTURE, with the two remaining populations, KNK (IL) and VIGO (IN), showing around 25% contribution from the eastern genetic cluster (Fig. 3). Over the whole range, average admixture is about 9% higher in agricultural populations than in non-agricultural populations: however, the average admixture of weedy populations is still less than 25%. Additionally, many individual non-weed populations, especially those in Indiana, have levels of admixture equal to or greater than those observed in agricultural habitats in STRUCTURE; the result is also seen in the heterozygosity comparisons in Table 2. The geographical component of the observed admixture suggests that it was simply an inevitable consequence of *A. tuberculatus* var. *rudis* migrating eastward.

We intensively sampled the edge of the range of agricultural waterhemp in Ohio, in order to examine evidence for weedy admixture at a finer geographic scale in this region with recent contact between varieties and presumably, little historical gene flow. According to STRUCTURE analysis of the total range dataset, admixture is prevalent in 10 out of the 12 sampled Ohio populations (including six of the eight non-agricultural populations) (Fig. 3). However, when only the Ohio populations are analyzed in STRUCTURE, admixture was 25% or less in the four agricultural populations (CAN,

GTB, MC, and RT29), which were mainly assigned to the western genetic cluster (Fig. 5). In fact, average admixture was almost 20% lower within the 10 counties that constitute the "agricultural waterhemp region" of the state than in the remainder of Ohio. Together with the whole range results, this suggests that admixture is not a prerequisite for weediness in *A. tuberculatus*.

The population structure data suggest that the movement of *A. tuberculatus* var. *rudis* eastward almost completely replaced *A. tuberculatus* var. *tuberculatus* populations in natural environments in central Illinois (e.g., the natural populations RIP and KEY have very little signature of the eastern genetic cluster). Interestingly, the central part of Indiana (IND) is genetically more "eastern" than the western part of Ohio, potentially indicating that *A. tuberculatus var. rudis* might have been secondarily introduced to Ohio from farther west. The populations in the "agricultural waterhemp region" of Ohio are either strongly admixed or primarily western genetically in the STRUCTURE analyses (Figs. 3, 5), and three of these populations are genetically western according to the BAPS Voroni diagram (Fig. 4b). The range boundary of agricultural waterhemp is around Columbus, Ohio (J. Stachler, NDSU, pers. comm.), and natural populations were not sampled in the eastern half of the state. The inclusion of additional eastern populations could confirm the inference that *A. tuberculatus* var. *rudis* genetic material hits a range boundary in central Ohio.

The western variety of waterhemp may also have northern and southern geographic boundaries in the areas it has invaded: the northern Illinois populations PEK and KNK and the northern Indiana populations KANK and WAB show more evidence of eastern genetic ancestry than populations in the middle of those states. The southern

populations along the Ohio River in Indiana and Ohio are almost entirely composed of the eastern genetic cluster, and the southern population BTL and the northern populations OTT, MAU, and PTC are eastern according to BAPS or admixed according to STRUCTURE.

Altogether, these patterns of genetic clustering point to a geographical invasion of *A. tuberculatus* var. *rudis* almost directly eastward through the primary agricultural regions of the eastern states, facilitated by introduction first in crop fields (as observed by Sauer, 1957). Waterhemp weed seeds are extensively moved around by farm equipment, which is often shared between farms and transported long distances (Patrick Tranel, University of IL, pers. comm.). With the evolution of resistance to multiple herbicide classes in the species, the spread of *A. tuberculatus* var. *rudis* throughout the Midwest became highly probable. The reasons for the possible northern, southern, and eastern geographical boundaries to the invasion deserve further study: the Ohio boundary may involve soil substrate (which changes abruptly in the middle of the state, in Licking County, east of Columbus (ODNR, 2014)), and the northern and southern boundaries are more likely to involve differences in climate and topography.

The origin of weedy waterhemp—Returning to the major question of the study — how the weedy form of *A. tuberculatus* arose — introgression between the two varieties is not supported as a causative factor in this study. The most likely scenario based on our results is that the weed form is simply *A. tuberculatus* var. *rudis*, which was already genetically and phenotypically suited to agricultural environments. When Mississippi Valley environments became increasingly dominated by agriculture in the 20th century, due to large-scale mechanized farming and the channeling of rivers for the greater

528 agricultural availability of floodplain habitats (Ghersa et al., 1994), *A. tuberculatus* var.
529 *rudis* was already well-suited to coexist and compete with crops in these new
530 environments. Later in the 20th century, the further expansion of waterhemp as a weed
531 was facilitated by the widespread adoption of no-till agriculture and herbicide-based
532 weed control (Costea et al., 2005). The idea that *A. tuberculatus* var. *rudis* was already
533 "weedy" and might not have required genetic changes to be successful in agricultural
534 ecosystems is supported by Sauer's description of the taxon, in which he states that in
535 contrast to *A. tuberculatus* var. *tuberculatus*, var. *rudis* has "very definite weedy
536 tendencies," and one-third of the herbarium collections of the species are from artificial,
537 anthropogenically-disturbed habitats (Sauer, 1955). Further evidence for the agricultural
538 inclinations of the western variety of *A. tuberculatus* was discovered in a separate
539 common garden experiment (K. Waselkov, unpublished data).

540 In the few other population genetic studies of agricultural invasion by wild plants,
541 where gene flow from cultivated plants is unlikely, a key conclusion is often exaptation
542 or "preadaptation" of these species to agricultural environments (Vigueira et al., 2013).
543 For example, weedy *Helianthus annuus* populations appear to have arisen multiple times
544 without genetic bottlenecks (Kane and Rieseberg, 2008), and an important parasite of
545 cereal crops, *Striga hermonthica*, shows no evidence of host specificity across Ethiopia,
546 with genetic structure influenced instead by geography (Welsh and Mohamed, 2011).
547 Exceptions to this trend include crop mimicry, in which weeds are strongly selected to
548 resemble the co-occurring crop during one ore more life history stages (Barrett, 1983),
549 and also herbicide avoidance and/or resistance, although these adaptive changes would

require enough preliminary tolerance to agricultural practices to enable the wild plants to occur in crop fields, before selection could act.

Our interest in *Amaranthus tuberculatus* stemmed not only from its "wild-to-weed" path to agricultural invasion, but also from the opportunity to trace this evolutionary transition in a native plant. Given the often close proximity of agricultural fields and natural riverbank habitats, it is perhaps surprising that any genetic signature of the two varieties, let alone a genetic signature of the eastward invasion of *A. tuberculatus* var. *rudis*, still exists. Indeed, our genetic diversity results support the idea that gene flow homogenizes neutral genetic variation over large areas of the species range (Thinglum, 2010). This is evident in the isolation by distance analyses, which show no isolation by distance at scales smaller than groupings of three states or provinces, and the low overall Fst in the total range dataset. These results are also consistent with the species' biology, as it is locally abundant, obligately outcrossing, and wind pollinated, and thus probably has very large effective population sizes and very high effective recombination across the genome (Thinglum et al., 2011).

The lack of complete genetic homogenization of waterhemp after secondary contact between the two varieties might suggest some degree of reproductive isolation between them: however, greenhouse experiments have found no evidence for pre- or postzygotic reproductive barriers between the two varieties (Murray 1940). Furthermore, pollen of *A. tuberculatus* is viable for up to 120 hours, allowing for long-distance dispersal, although most pollen fertilizes plants within 50 meters in field trials (Liu et al., 2012). There is some anecdotal evidence that prezygotic isolation may be occurring in the wild: the timing of reproduction appears to differ between riverbanks and crop fields,

573 with agricultural waterhemp senescing while nearby riverbank waterhemp is just
574 beginning to flower (K. Waselkov, pers. obs.). Further studies of observed dispersal and
575 realized gene flow between the different varieties and habitats in this species could shed
576 light on the apparent contradiction between our results and the species' biology.

577 Our results also suggest that finer-scale geographic sampling as well as more
578 genetic markers and/or candidate genes would be needed for studies attempting to detect
579 the genetic basis of agricultural adaptation in *A. tuberculatus*. Various herbicide
580 resistance genes are already known to be strongly favored in agricultural populations of
581 *A. tuberculatus* (Tranel and Trucco, 2009); population genetic studies with denser
582 genomic coverage could reveal other specific genomic regions showing adaptive
583 differentiation between agricultural and natural environments (e.g., Loh et al., 2008;
584 Bouchet et al., 2012). Gene expression differences were detected using a cDNA
585 microarray between nearby wild and weedy populations of sunflowers, suggesting both
586 agricultural and local adaptation (Lai et al, 2008); however, a European study of the
587 rapidly expanding, wind-pollinated weed blackgrass (*Alopecurus myosuroides*) found
588 that even though agricultural populations experienced strong selection from herbicide
589 application, this did not modify their genetic structure at 116 AFLPs distributed across
590 the genome (Menchari et al., 2007).

591 ***Conclusions***—In this study, we have built on the observations of Sauer (1957)
592 and Pratt and Clark (2001) to present a new hypothesis about the origin of the agricultural
593 weed form of *A. tuberculatus*. Weedy *A. tuberculatus* var. *rudis* appears to have blazed a
594 path from west to east through the central Midwestern states without significant genetic
595 change at neutral markers. Its surprisingly limited admixture with the native conspecific

eastern variety is intriguing, and should be investigated further. Evolution in response to agricultural management practices is ongoing in this species, as exemplified by its continual adaptation to new herbicides (e.g., Hausman et al., 2011), and current range boundaries may shift in response to evolution or land use changes.

This research shows that the most common mode of weed origination, the expansion of the niches of wild plants to include agricultural environments, can proceed via quite different evolutionary trajectories from weeds related to domesticated plants. Future evolutionary investigations of the understudied wild-to-weed category of weed origins have the potential to show how general or specific the patterns found in this paper are. Generalizations from these studies would be interesting from the perspective of both basic evolutionary biology and weed control.

619 LITERATURE CITED
620
621 AONO, M., S. WAKIYAMA, M. NAGATSU, N. NAKAJIMA, M. TAMAOKI, A. KUBO, AND H.
622 SAJI. 2006. Detection of feral transgenic oilseed rape with multiple-herbicide
623 resistance in Japan. *Environmental Biosafety Research* 5: 77-87.
624
625 BARRETT, S. C. H. 1983. Crop mimicry in weeds. *Economic Botany* 37: 255-282.
626
627 BOUCHET, S., D. POT, M. DEU, J.-F. RAMI, C. BILLOT, X. PERRIER, R. RIVALLAN, ET AL.
628 2012. Genetic structure, linkage disequilibrium and signature of selection in
629 sorghum: lessons from physically anchored DArT markers. *PLoS ONE* 7:
630 e33470.
631
632 BROWNSTEIN, M. J., J. D. CARPTEN, AND J. R. SMITH. 1996. Modulation of non-templated
633 nucleotide addition by *Taq* DNA polymerase: primer modifications that facilitate
634 genotyping. *BioTechniques* 20: 1004-1010.
635
636 BUHLER, D. D., R. G. HARTLZER, AND F. FORCELLA. 1997. Implications of weed seedbank
637 dynamics to weed management. *Weed Science* 45: 329-336.
638
639 BURGER, J. C., S. LEE, AND N. C. ELLSTRAND. 2006. Origin and genetic structure of feral
640 rye in the western United States. *Molecular Ecology* 15: 2527-2539.
641
642 CAMPBELL, L. G., A. A. SNOW, AND C. E. RIDLEY. 2006. Weed evolution after crop gene
643 introgression: greater survival and fecundity of hybrids in a new environment.
644 *Ecology Letters* 9: 1198-1209.
645
646 CLEMENTS, D. R., A. DITOMMASO, N. JORDAN, B. D. BOOTH, J. CARDINA, D. DOOHAN, C.
647 L. MOHLER, ET AL. 2004. Adaptibility of plants invading North American
648 cropland. *Agriculture, Ecosystems and Environment* 104: 379-398.
649
650 CORANDER, J., P. WALDMANN, AND M. J. SILLANPÄÄ. 2003. Bayesian analysis of genetic
651 differentiation between populations. *Genetics* 163: 367-374.
652
653 CORANDER, J., J. SIRÉN, AND E. ARJAS. 2008. Bayesian spatial modeling of genetic
654 population structure. *Computational Statistics* 23: 111-129.
655
656 COSTEA, M., AND F. J. TARDIF. 2003. Conspectus and notes on the genus *Amaranthus* in
657 Canada. *Rhodora* 105: 260-281.
658
659 COSTEA, M., S. E. WEAVER, AND F. TARDIF. 2005. The biology of invasive alien plants in
660 Canada. 3. *Amaranthus tuberculatus* (Moq.) Sauer var. *rudis* (Sauer) Costea &
661 Tardif. *Canadian Journal of Plant Science* 85: 507-522.
662
663 DE WET, J. M. J., AND J. R. HARLAN. 1975. Weeds and domesticates: evolution in the
664 man-made habitat. *Economic Botany* 29: 99-107.

EARL, D. A., AND VONHOLDT, B. M. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359-361.

ERSTS, P. J. [Internet] Geographic Distance Matrix Generator (version 1.2.3). American Museum of Natural History, Center for Biodiversity and Conservation. Available at http://biodiversityinformatics.amnh.org/open_source/gdmg. Accessed on 2013-6-2.

EVANNO, G., S. REGNAUT, AND J. GOUDET. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611-2620.

FÉNART, S., J-F. ARNAUD, I. DE CAUWER, AND J. CUGUEN. 2008. Nuclear and cytoplasmic genetic diversity in weed beet and sugar beet accessions compared to wild relatives: new insights into the genetic relationships within the *Beta vulgaris* complex species. *Theoretical and Applied Genetics* 116: 1063-1077.

FRANÇOIS, O., AND E. DURAND. 2010. Spatially explicity Bayesian clustering models in population genetics. *Molecular Ecology Resources* 10: 773-784.

GHERSA, C. M., M. L. ROUSH, S. R. RADOSEVICH, AND S. M. CORDRAY. 1994. Coevolution of agroecosystems and weed management. *BioScience* 44: 85-94.

GOUDET, J. 1995. FSTAT Version 1.2: a computer program to calculate F-statistics. *Journal of Heredity* 86: 485-486. FSTAT 2.9.3.2 software available from: http://www2.unil.ch/popgen/softwares/fstat.htm

HAGER, A., L. WAX, W. SIMMONS, AND C. SPRAGUE. 2000. Waterhemp management in Illinois agronomic crops. *In* the University of Illinois Extension [eds.], 2000 Illinois Agricultural Pest Management Handbook, 91-100. The University of Illinois, Urbana-Champaign, Illinois, USA.

HAUSMAN, N. E., S. SINGH, P. J. TRANEL, D. E. RIECHERS, S. S. KAUNDUN, N. D. POLGE, D. A. THOMAS, AND A. G. HAGER. 2011. Resistance to HPPD-inhibiting herbicides in a population of waterhemp (*Amaranthus tuberculatus*) from Illinois, United States. *Pest Management Science* 67: 258-261.

JAKOBSSON, M., AND N. A. ROSENBERG. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23: 1801-1806.

KANE, N. C., AND L. H. RIESEBERG. 2008. Genetics and evolution of weedy *Helianthus annuus* populations: adaptation of an agricultural weed. *Molecular Ecology* 17: 384-394.

LAI, Z., N. C. KANE, Y. ZOU, AND L. H. RIESEBERG. 2008. Natural variation in gene expression between wild and weedy populations of *Helianthus annuus*. *Genetics* 179: 1881-1890.

LAMBRINOS, J. G. 2004. How interactions between ecology and evolution influence contemporary invasion dynamics. *Ecology* 85: 2061-2070.

LEE, R. M., J. THIMMAPURAM, K. A. THINGLUM, G. GONG, A. G. HERNANDEZ, C. L. WRIGHT, R. W. KIM, M. A. MIKEL, AND P. J. TRANEL. 2009. Sampling the waterhemp (*Amaranthus tuberculatus*) genome using pyrosequencing technology. *Weed Science* 57: 463-469.

LOH, Y.-H. E., L. S. KATZ, M. C. MIMS, T. D. KOCHER, S. V. YI, AND J. T. STREELMAN. 2008. Comparative analysis reveals signatures of differentiation amid genomic polymorphism in Lake Malawi cichlids. *Genome Biology* 9: R113.

LONDO, J. P., AND B. A. SCHAAL. 2007. Origins and population genetics of weedy red rice in the USA. *Molecular Ecology* 16: 4523-4535.

LIU, J., A. S. DAVIS, AND P. J. TRANEL. 2012. Pollen biology and dispersal dynamics in waterhemp (*Amaranthus tuberculatus*). *Weed Science* 60: 416-422.

MACK, R. N., D. SIMBERLOFF, W. M. LONSDALE, H. EVANS, M. CLOUT, AND F. A. BAZZAZ. 2000. Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications* 10: 680-710.

MENCHARI, Y., C. DÉLYE, AND V. LE CORRE. 2007. Genetic variation and population structure in black-grass (*Alopecurus myosuroides* Huds.), a successful, herbicide-resistant, annual grass weed of winter cereal fields. *Molecular Ecology* 16: 3161-3172.

MORRELL, P. L., T. D. WILLIAMS-COPLIN, A. L. LATTU, J. E. BOWERS, J. M. CHANDLER, AND A. H. PATERSON. 2005. Crop-to-weed introgression has impacted allelic composition of johnsongrass populations with and without recent exposure to cultivated sorghum. *Molecular Ecology* 14: 2143-2154.

MOSYAKIN, S., AND K. R. ROBERTSON. 2003. *Amaranthus*. *In* Flora of North America Editorial Committee [eds.], Flora of North America North of Mexico, Vol. 4. Magnoliophyta: Caryophyllidae, Pt. 1. Oxford University Press, New York, New York, USA.

MULLER, M.-H., M. LATREILLE, AND C. TOLLON. 2010. The origin and evolution of a recent agricultural weed: population genetic diversity of weedy populations of sunflower (*Helianthus annuus* L.) in Spain and France. *Evolutionary Applications* 4: 499-514.

MURRAY, J.M. 1940. The genetics of sex determination in the family Amaranthaceae. *Genetics* 25: 409–431.

ODNR. 2014. Soil regions of Ohio, Brochure 1. Division of Soil and Water Resources, Ohio Department of Natural Resources, Columbus, Ohio, U.S.A. Available at http://www.dnr.state.oh.us/tabid/9073/Default.aspx

OLSEN, K. M., A. L. CAICEDO, AND Y. JIA. 2007. Evolutionary genomics of weedy rice in the USA. *Journal of Integrative Plant Biology* 49: 811-816.

OWEN, M. D. K. 2008. Weed species shifts in glyphosate-resistant crops. *Pest Management Science* 64: 377-387.

PIMENTEL, D., R. ZUNIGA, AND D. MORRISON. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics* 52: 273-288.

PRATT, D. B., AND L. G. CLARK. 2001. *Amaranthus rudis* and *A. tuberculatus*, one species or two? *Journal of the Torrey Botanical Society* 128: 282-296.

PRITCHARD, J. K., M. STEPHENS, AND P. DONNELLY. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.

REFSELL, D. E., AND R. G. HARTZLER. 2009. Effect of tillage on common waterhemp (*Amaranthus rudis*) emergence and vertical distribution of seed in the soil. *Weed Technology* 23: 129-133.

ROSENBERG, N. A. 2004. *Distruct*: a program for the graphical display of population structure. *Molecular Ecology Notes* 4: 137-138.

ROUSSET, F. 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources* 8: 103-106.

SAKAI, A. K., F. W. ALLENDORF, J. S. HOLT, D. M. LODGE, J. MOLOFSKY, K. A. WITH, S. BAUGHMAN, ET AL. 2001. The population biology of invasive species. *Annual Review of Ecology and Systematics* 32: 305-332.

SAUER, J. D. 1955. Revision of the dioecious amaranths. *Madroño* 13: 5-46.

SAUER, J. D. 1957. Recent migration and evolution of the dioecious amaranths. *Evolution* 11: 11-31.

SAUER, J. D. 1972. The dioecious amaranths: a new species name and major range extensions. *Madroño* 21: 427-434.

SCHROEDER, S. 2003. SSR Finder. Maize Mapping Project, University of Missouri, Columbia, MO, USA.

SCHUELKE, M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18: 233-234.

STECKEL, L. E. 2007. The dioecious *Amaranthus* spp.: here to stay. *Weed Technology* 21: 567-570.

THINGLUM, K. A. 2010. Population genetics of herbicide resistance in waterhemp. Masters Thesis, University of Illinois at Urbana-Champaign. Available through the Illinois Digital Environment for Access to Learning and Scholarship.

THINGLUM, K. A., C. W. RIGGINS, A. S. DAVIS, K. W. BRADLEY, K. AL-KHATIB, AND P. J. TRANEL. 2011. Wide distribution of the waterhemp (*Amaranthus tuberculatus*) ΔG210 *PPX2* mutation, which confers resistance to PPO-inhibiting herbicides. *Weed Science* 59: 22-27.

TRANEL, P. J., AND F. TRUCCO. 2009. 21st century weed science: a call for *Amaranthus* genomics. *In* C. N. Stewart Jr. [ed.], Weedy and Invasive Plant Genomics, 53-81. Blackwell Publishing, Ames, IA, USA.

VAN OOSTERHOUT, C., W. F. HUTCHINSON, D. P. M. WILLS, AND P. SHIPLEY. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4: 535-538.

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.

WARD, S. M., J. F. GASKIN, AND L. M. WILSON. 2008. Ecological genetics of plant invasion: what do we know? *Invasive Plant Science and Management* 1: 98-109.

WARWICK, S. I., M.-J. SIMARD, A. LÉGÈRE, H. J. BECKIE, L. BRAUN, B. ZHU, P. MASON, G. SÉGUIN-SWARTZ, AND C. N. STEWART. 2003. Hybridization between transgenic *Brassica napus* L. and its wild relatives: *Brassica rapa* L., *Raphanus raphanistrum* L., *Sinapis arvensis* L., and *Erucastrum gallicum* (Willd.) O.E. Schulz. *Theoretical and Applied Genetics* 107: 528-539.

WELSH, A. B., AND K. I. MOHAMED. 2011. Genetic diversity of *Striga hermonthica* populations in Ethiopia: evaluating the role of geography and host specificity in shaping population structure. *International Journal of Plant Sciences* 172: 773-782.

YEH, F.C., R.-C. YANG, T. BOYLE, Z.-H. YE, AND J.X. MAO. 1997. POPGENE: the user-friendly shareware for population genetic analysis. Molecular Biology and Biotechnology Centre, University of Alberta, Canada. Available at http://www.ualberta.ca/~fyeh/

TABLES

Table 1. Genotyped populations of *Amaranthus tuberculatus*: name, U.S. state/Canadian province, geographical region of population (from Figure 1: west = historical range of *A. tuberculatus* var. *rudis*, east = historical range of *A. tuberculatus* var. *tuberculatus*, overlap = historical range of both varieties), habitat of population (agricultural or natural), number of individuals genotyped (N), and voucher specimen fruit dehiscence (whether the ripe utricle opens or not).

Name	State/ Province	Geographical region	Habitat	N	Voucher fruit dehiscence
SaltR	OK	west	riverbank	10	dehiscent
CHE	KS	west	lake shore	10	too young to tell
TCL	KS	west	lake shore	10	dehiscent
TR	NE	west	riverbank	9	dehiscent
MSH	MO	west	soy field	10	dehiscent
GASC	MO	west	riverbank	10	too young to tell
JCK	AR	west	riverbank	9	dehiscent
WSS	MO	west	sunflower field	7	dehiscent
WSR	MO	west	riverbank	10	indehiscent
AAF	MO	west	soy field	10	dehiscent
RIP	IL	west	riverbank	10	indehiscent
GTP	MO	west	riverbank	10	dehiscent
EMN	MO	west	riverbank	8	dehiscent
PEK	IL	overlap	riverbank	10	indehiscent
KEY	IL	overlap	lake shore	10	too young to tell
KEYC	IL	overlap	soy field	10	dehiscent
KNK	IL	overlap	soy field	10	indehiscent
VIGO	IN	overlap	soy field	10	indehiscent
KANK	IN	east	riverbank	10	indehiscent
IND	IN	east	riverbank	10	indehiscent
WAB	IN	east	riverbank	10	indehiscent
AUR	IN	east	riverbank	10	too young to tell
DMD	MI	east	riverbank	9	indehiscent
RT29	OH	east	corn field	10	dehiscent
BTL	OH	east	riverbank	9	dehiscent
GTB	OH	east	soy field	10	indehiscent
STW	OH	east	riverbank	10	indehiscent
NEV	OH	east	riverbank	10	indehiscent
OTT	OH	east	riverbank	10	indehiscent
MAU	OH	east	riverbank	10	indehiscent
MC	OH	east	soy field	10	indehiscent
PCL	OH	east	lake shore	10	indehiscent
CAN	OH	east	soy field	10	indehiscent
SCIO	OH	east	lake shore	10	dehiscent
PTC	OH	east	lake shore	9	indehiscent
SCF	ON	east	lake shore	10	too young to tell
DEL	ON	east	riverbank	9	indehiscent
YORK	ON	east	riverbank	10	indehiscent

Table 2. Population genetic statistics for each population summarized over all loci. Na = number of alleles, Ne = effective number of alleles (estimated reciprocal of homozygosity), Ho = observed heterozygosity, He = expected heterozygosity. Populations with higher observed than expected heterozygosity (potentially reflecting admixture) are in bold font.

Population	Mean Na	Mean Ne	Mean Ho	Mean He
TR	5.1	3.22	0.57	0.65
CHE	4.9	3.48	0.57	0.60
TCL	5.2	3.05	0.58	0.59
SaltR	5.2	3.50	0.65	0.68
GASC	4.9	3.09	0.54	0.54
MSH	4.8	3.07	0.50	0.55
AAF	4.7	3.39	0.60	0.61
WSR	4.9	3.39	0.57	0.61
GTP	5.1	3.52	0.59	0.62
WSS	4.4	3.44	0.63	0.64
EMN	4.6	3.21	0.49	0.55
JCK	5.2	3.70	0.51	0.62
PEK	4.8	3.31	**0.59**	**0.58**
RIP	5.1	3.12	0.54	0.57
KEY	5.0	3.27	0.55	0.60
KEYC	4.7	3.17	0.51	0.56
KNK	5.3	3.18	0.55	0.60
VIGO	4.3	2.56	**0.69**	**0.55**
KANK	4.8	3.22	**0.61**	**0.60**
WAB	5.3	3.44	0.63	0.64
IND	5.6	4.05	0.59	0.66
AUR	4.4	3.01	0.46	0.54
BTL	4.7	3.00	0.51	0.58
PCL	3.9	2.86	0.49	0.50
NEV	3.8	2.64	**0.52**	**0.49**
STW	5.0	3.55	0.57	0.60
CAN	3.3	2.34	**0.58**	**0.55**
SCIO	4.5	3.27	**0.64**	**0.63**
GTB	4.5	2.74	0.47	0.53
MC	3.7	2.44	**0.62**	**0.58**
RT29	5.0	3.64	0.60	0.64
OTT	4.2	2.83	0.49	0.59
MAU	4.5	3.08	0.54	0.60
PTC	4.6	2.96	**0.59**	**0.58**
DMD	4.2	2.91	0.50	0.55
DEL	3.0	2.01	0.46	0.46
SCF	4.7	3.47	0.60	0.63
YORK	3.5	2.01	0.38	0.42
All Populations	4.6	3.11	0.56	0.58

862 FIGURE LEGENDS
863
864 Figure 1. Locations of 38 genotyped populations of *A. tuberculatus* from across the
865 species range. Geographic coordinates were plotted in ArcGIS, and each population is
866 shown as an orange circle, with the population name in orange text (corresponding to
867 Table 1). The green line outlines the geographical region corresponding to the historical
868 range of *A. tuberculatus* var. *rudis*, and the purple line outlines the historical range of *A.*
869 *tuberculatus* var. *tuberculatus* (adapted from Sauer, 1957). The red dashed trapezoid
870 outlines the "agricultural waterhemp region" of Ohio. Map sources: ESRI, DeLorme,
871 USGS, NOAA.
872
873 Figure 2. Plot of pairwise genetic distances (Fst/(1-Fst)) versus pairwise geographic
874 distances (ln(kilometers)) for the 38 *A. tuberculatus* populations genotyped over the
875 entire species range, showing isolation by distance. The regression has been constrained
876 to go through the origin.
877
878 Figure 3. STRUCTURE population assignment of individuals for K=2, for the 38
879 genotyped populations from across the species range. The “western” genetic cluster is in
880 green and the “eastern” genetic cluster is in purple. Population names are shown below
881 the bar graph, and the arrow below the graph shows the geographical organization of the
882 populations. A red asterisk above a population indicates it was found in an agricultural
883 habitat.
884

885 Figure 4. BAPS graphical output for K=3, for the 38-population species' range dataset.
886 The "western" genetic cluster is in green and the "eastern" genetic cluster is in blue,
887 while the third cluster (the MC population) is in red. A. The admixture bar graph for
888 K=3. Population names are shown below the population clusters, and the organization of
889 populations geographically is the same as in Figure 3. B. The Voroni tessellation
890 diagram for K=3, labeled with population names.
891
892 Figure 5. STRUCTURE population assignment of individuals for K=2, for the 12-
893 population Ohio dataset. The "western" genetic cluster is in green and the "eastern"
894 genetic cluster is in purple. Population names are shown below the bar graph, a red
895 asterisk above a population indicates it was found in an agricultural habitat, and a black
896 asterisk indicates a population found in the "agricultural waterhemp region" but in a non-
897 agricultural habitat.
898
899
900 SUPPORTING INFORMATION
901 Table S1. Genotyped populations of *Amaranthus tuberculatus*, with name, U.S.
902 state/Canadian province, voucher information, and geographical coordinates.
903
904 Table S2. Microsatellite loci forward (F) and reverse (R) primers, repeat motif, dye label
905 (used in multiplexed reactions), size range, and primer source. Sources are Lee et al.
906 (2009) and the Tranel lab at the University of Illinois-Urbana/Champaign.
907

Table S3. Population genetic statistics for each locus summarized over all populations. Na = number of alleles, Ne = effective number of alleles (estimated reciprocal of homozygosity), Ho = observed heterozygosity, He = expected heterozygosity.

Figure S1. Locations of 12 genotyped populations of *A. tuberculatus* from Ohio. Geographic coordinates were plotted in ArcGIS, and each population is shown as an orange circle, with the population name in orange text (corresponding to Table 1). The red dashed trapezoid outlines the "agricultural waterhemp region" of the state (J. Stachler, NDSU, pers. comm.). Map sources: ESRI, DeLorme, USGS, NOAA.

Figure S2. Plots of (A) mean LnP(D) (the natural log of the probability of the data from STRUCTURE output) with vertical lines showing standard deviation, and (B) ΔK (the second-order rate of change of K) versus K, for the 38 genotyped populations from across the species range, showing a highest ΔK of 2. Plots are from the STRUCTURE HARVESTER output.

Figure S3. Plots of (A) mean LnP(D) (the natural log of the probability of the data from STRUCTURE output) with vertical lines showing standard deviation, and (B) ΔK (the second-order rate of change of K) versus K, for the 12 genotyped populations from Ohio, showing a highest ΔK of 2, and a second peak at K=5. Plots are from the STRUCTURE HARVESTER output.

930 Figure S4. STRUCTURE population assignment of individuals for K=5 for the 12-
931 population Ohio dataset. Population names are shown below the bar graph, a red asterisk
932 above a population indicates it was found in an agricultural habitat, and a black asterisk
933 indicates a population found in the "agricultural waterhemp region" but in a non-
934 agricultural habitat.
935
936

Figures

Click here to download Figure: Figures for PopGen Manuscript.ppt

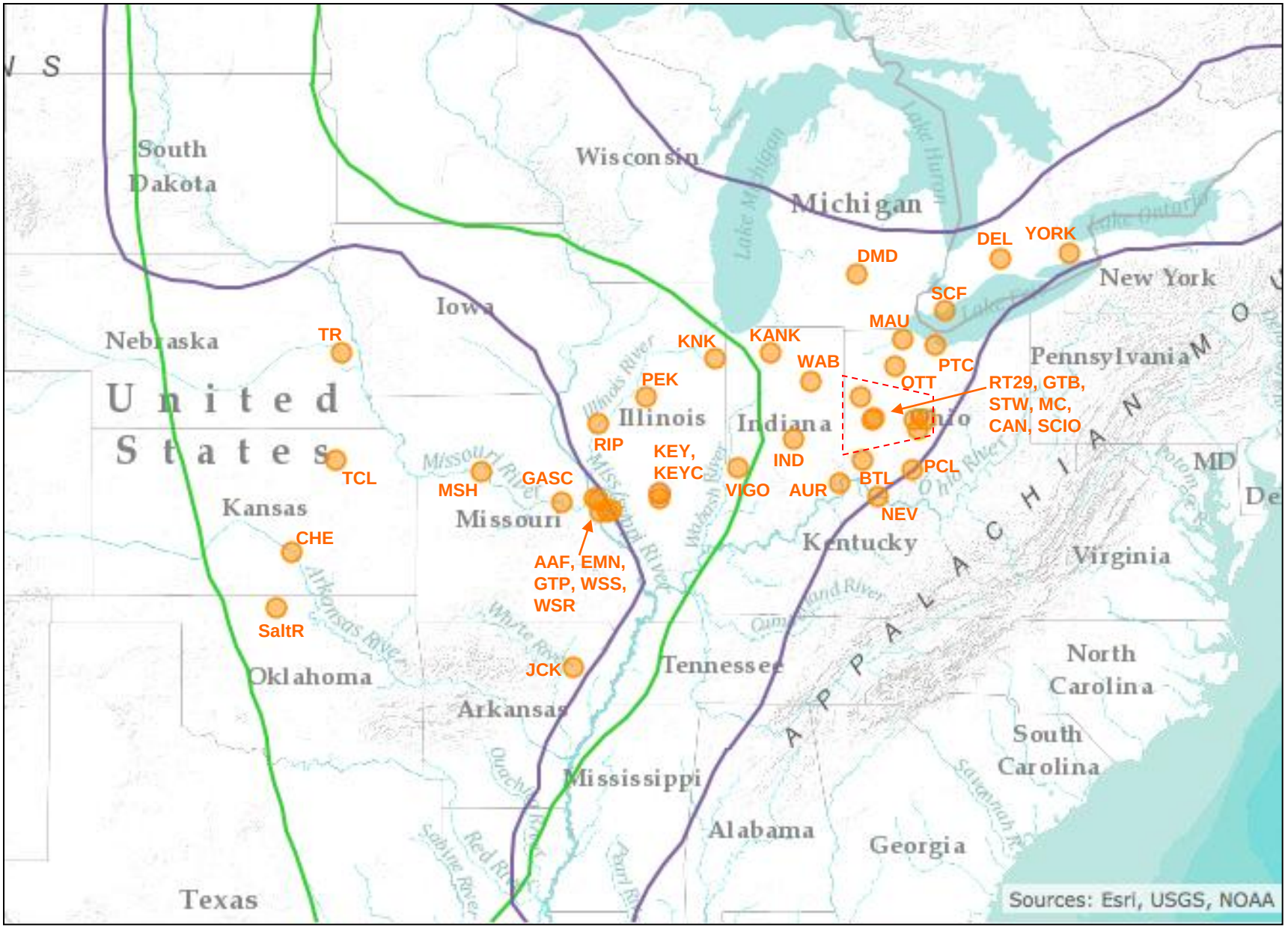


Figure 1.

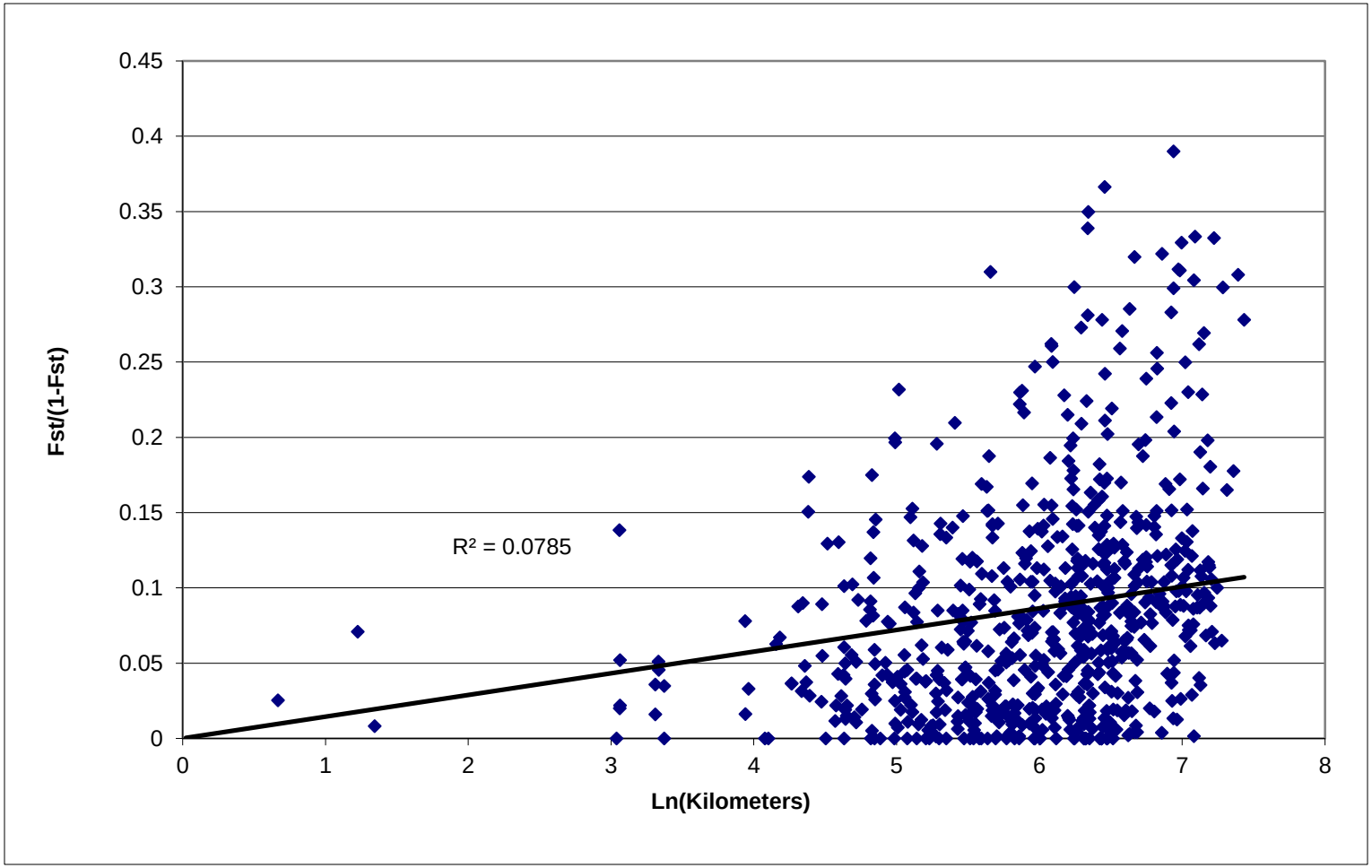

0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Fst/(1-Fst)
R² = 0.0785
0
1
2
3
4
5
6
7
8
Ln(Kilometers)

Figure 2.

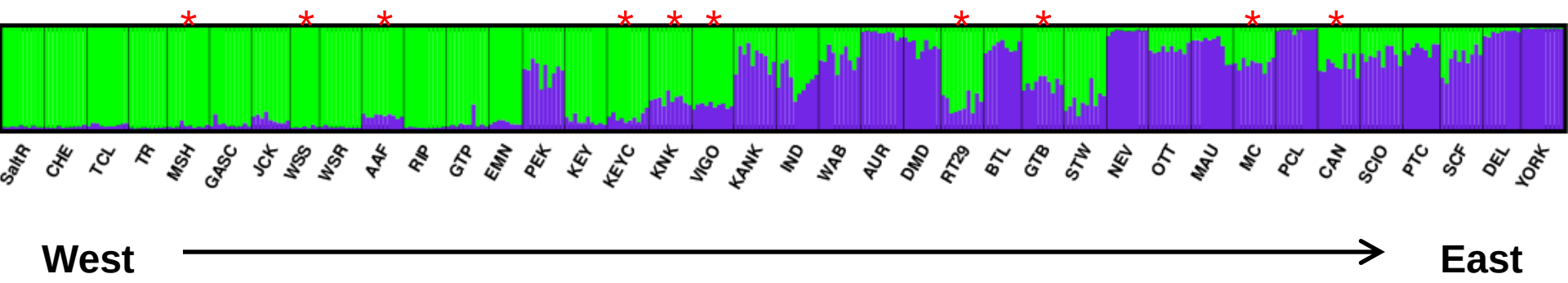

SaltR
CHE
TCL
TR
MSH
GASC
JCK
WSS
WSR
AAF
RIP
GTP
EMN
PEK
KEY
KEYC
KNK
VIGO
KANK
IND
WAB
AUR
DMD
RT29
BTL
GTB
STW
NEV
OTT
MAU
MC
PCL
CAN
SCIO
PTC
SCF
DEL
YORK
West
East

Figure 3.

A.

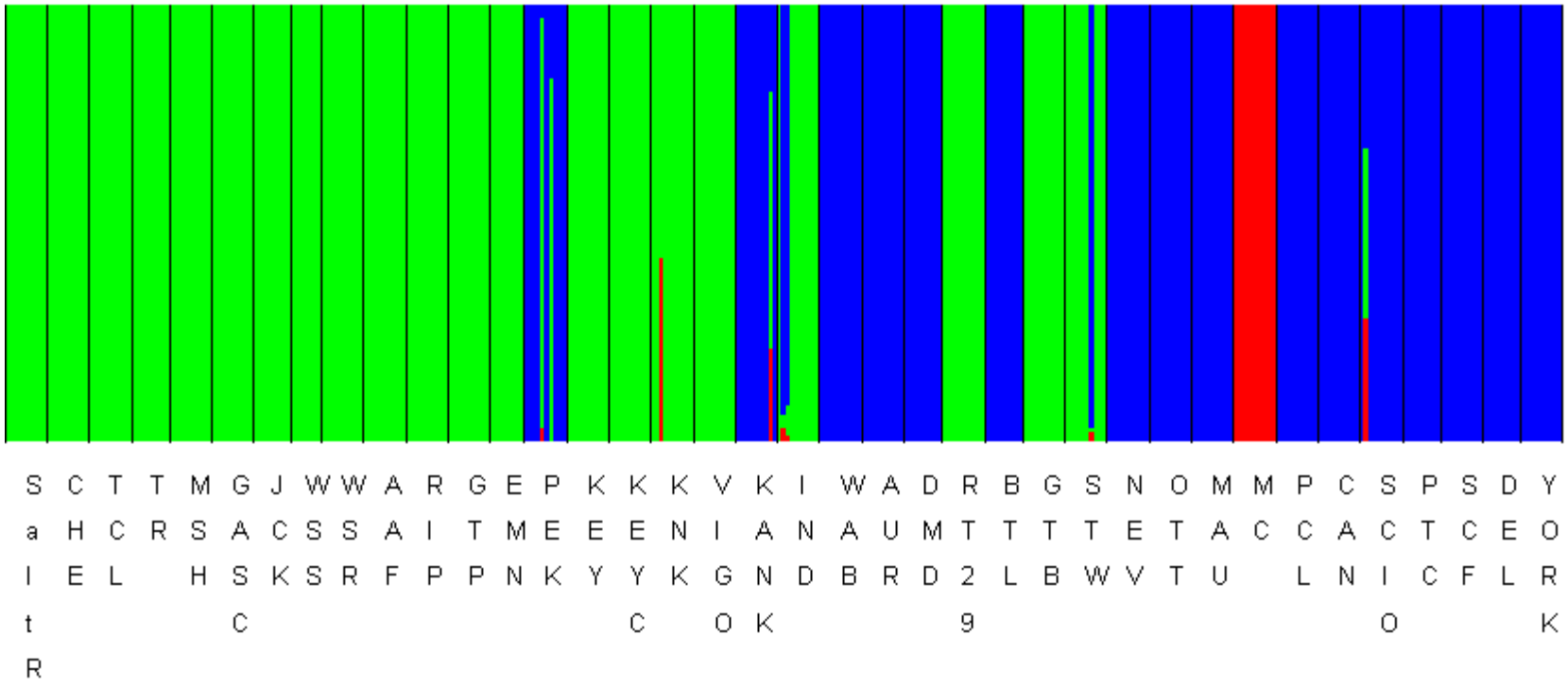

S C T T M G J W W A R G E P K K K V K I W A D R B G S N O M M P C S P S D Y
a H C R S A C S S A I T M E E E N I A N A U M T T T E T A C C A C T C E O
I E L H S K S R F P P N K Y Y K G N D B R D 2 L B W V T U L N I C F L R
t C C O K 9 O K
R

Figure 4a.

B.

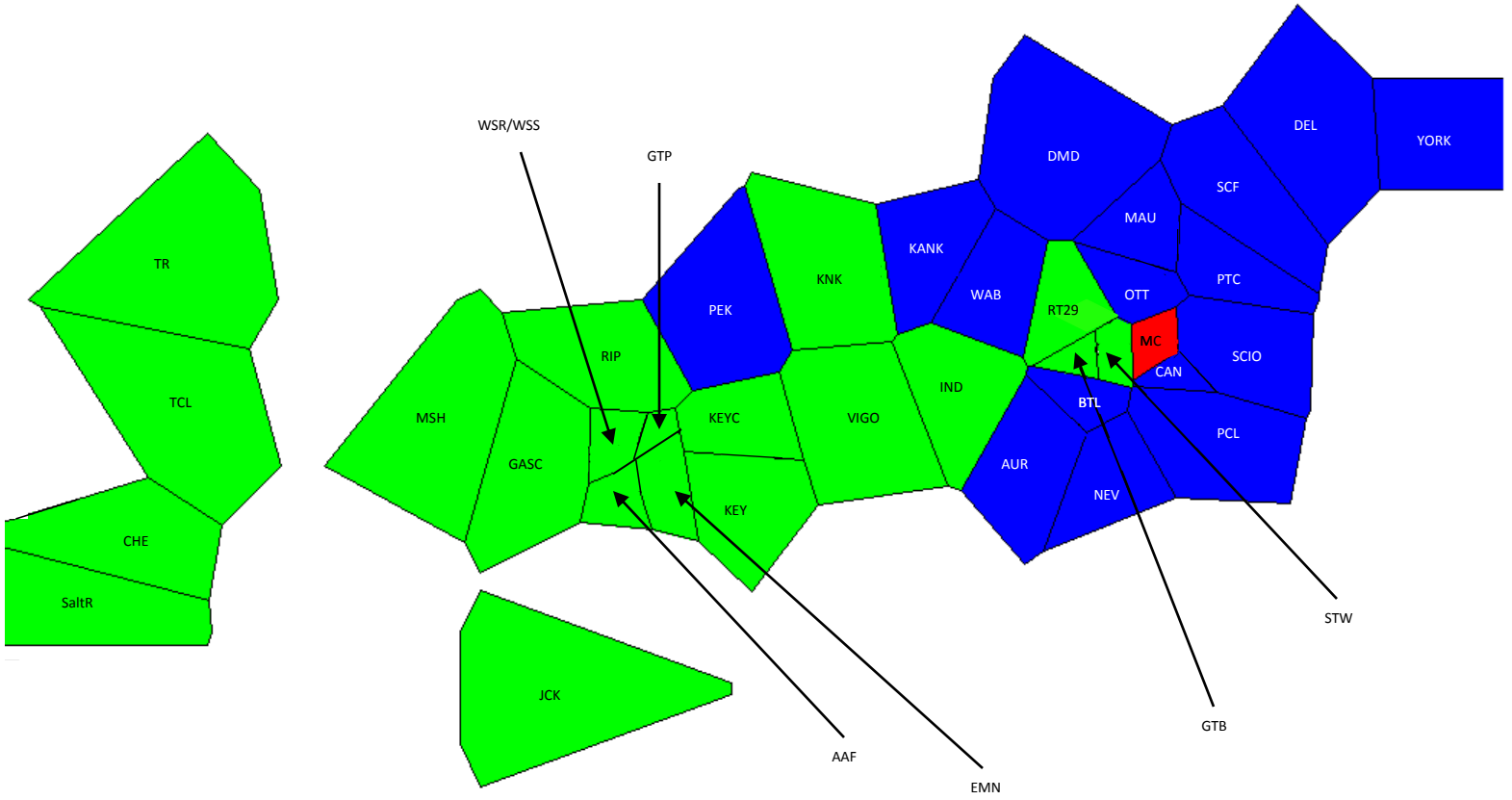

WSR/WSS
GTP
TR
TCL
CHE
SaltR
MSH
GASC
RIP
PEK
KNK
KANK
DMD
DEL
YORK
SCF
MAU
PTC
WAB
RT29
OTT
MC
CAN
SCIO
IND
KEYC
VIGO
BTL
PCL
AUR
NEV
KEY
JCK
AAF
EMN
GTB
STW

Figure 4b.

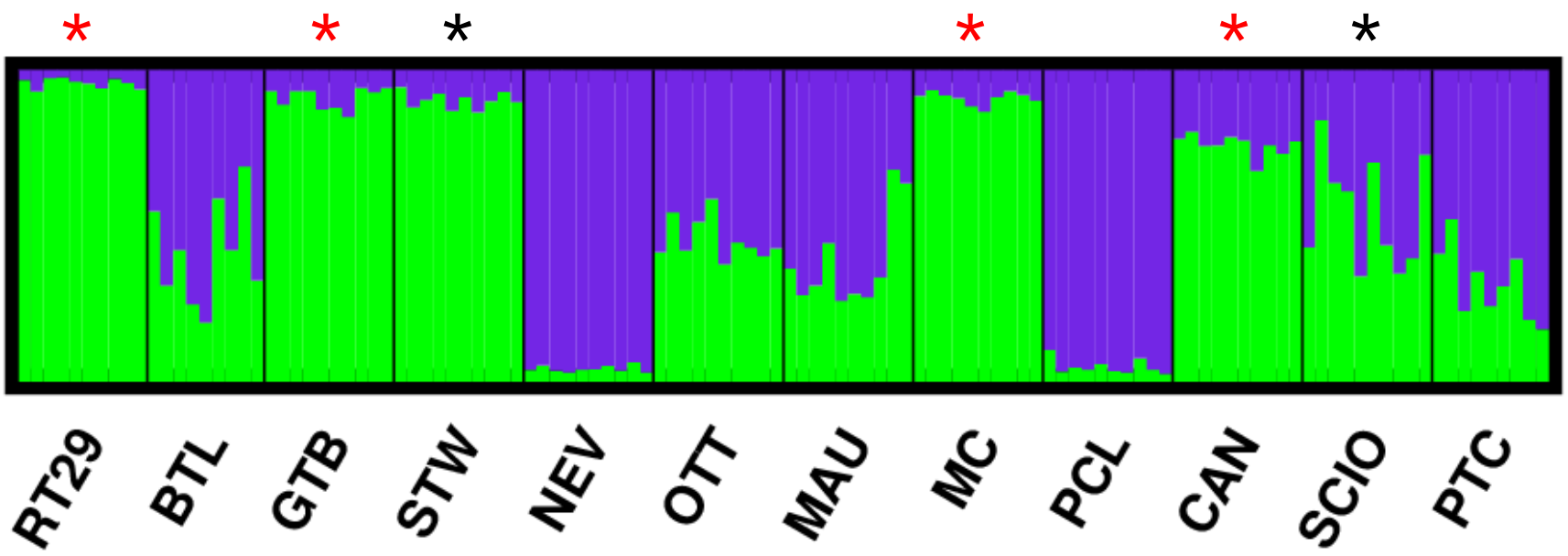


Figure 5.

Online Supplemental

Click here to download Online Supplemental: Supplementary Tables.docx

Table S1. Genotyped populations of *Amaranthus tuberculatus*, with name, U.S. state/Canadian province, voucher information, and geographical coordinates.

Name	State/ Province	Voucher Information: Collector, Collection #: Herbarium	Latitude	Longitude
SaltR	OK	K. Waselkov, #75: MO	36.771660	-98.038000
CHE	KS	K. Waselkov, #74: MO	37.744750	-97.783860
TCL	KS	K. Waselkov, #72: MO	39.439230	-96.710250
TR	NE	K. Waselkov, #68: MO	41.223310	-96.357610
MSH	MO	K. Waselkov, #138: KSC	39.168783	-93.289057
GASC	MO	K. Waselkov, #137: MO	38.668122	-91.556135
JCK	AR	K. Waselkov, #133: MO	35.642113	-91.319192
WSS	MO	K. Waselkov, #31: MO	38.656280	-90.736955
WSR	MO	K. Waselkov, #32: MO	38.656280	-90.736950
AAF	MO	K. Waselkov, #25: KSC	38.473250	-90.661016
RIP	IL	K. Waselkov, #58: MO	40.027434	-90.631546
GTP	MO	K. Waselkov, #24: MO	38.558930	-90.447360
EMN	MO	K. Waselkov, #26: KSC	38.545160	-90.433450
PEK	IL	K. Waselkov, #56: MO	40.574410	-89.655980
KEY	IL	K. Waselkov, #53: MO	38.733710	-89.275850
KEYC	IL	K. Waselkov, #142: KSC	38.768113	-89.273209
KNK	IL	K. Waselkov, #64: MO	41.160983	-87.627515
VIGO	IN	K. Waselkov, #119: MO	39.273930	-87.470000
KANK	IN	K. Waselkov, #115: MO	41.314810	-86.737550
IND	IN	K. Waselkov, #109: MO	39.783310	-86.189750
WAB	IN	K. Waselkov, #111: MO	40.790980	-85.820860
AUR	IN	K. Waselkov, #106: MO	39.056110	-84.898350
DMD	MI	K. Waselkov, #82: MO	42.645000	-84.649700
RT29	OH	K. Waselkov, #101: MO	40.545911	-84.634131
BTL	OH	K. Waselkov, #95: MO	39.427430	-84.540710
GTB	OH	K. Waselkov, #97: MO	40.120100	-84.398680
STW	OH	K. Waselkov, #98: MO	40.121630	-84.358660
NEV	OH	K. Waselkov, #92: MO	38.807630	-84.211710
OTT	OH	K. Waselkov, #91: MO	41.037830	-83.813490
MAU	OH	K. Waselkov, #90: MO	41.556350	-83.662410
MC	OH	K. Waselkov, #102: MO	40.155580	-83.455330
PCL	OH	K. Waselkov, #94: MO	39.268010	-83.388610
CAN	OH	K. Waselkov, #104: MO	39.985850	-83.339630
SCIO	OH	K. Waselkov, #103: MO	40.177450	-83.126400
PTC	OH	K. Waselkov, #89: MO	41.514500	-82.938430
SCF	ON	K. Waselkov, #84: MO	42.030950	-82.603850
DEL	ON	K. Waselkov, #86: MO	42.933750	-81.421060
YORK	ON	K. Waselkov, #88: MO	43.020700	-79.891050

Table S2. Microsatellite loci forward (F) and reverse (R) primers, repeat motif, dye label (used in multiplexed reactions), size range, and primer source. Sources are Lee et al. (2009) and the Tranel lab at the University of Illinois-Urbana/Champaign.

Locus name	Primer	Repeat	Dye label	Size range	Source
C1140	F: 5'-TTGAAGACGACGATCTTTCTGGAT	$(GAT)_{10}$	6FAM	113-181 bp	Tranel lab *A. tuberculatus* transcriptome data
	R: 5'-CCCCTCTGTACACCATAATCGAAC				
C4097	F: 5'-ATCATCTTCTGCTAAGGCTGTTGG	$(ACC)_8$	NED	164-179 bp	Tranel lab *A. tuberculatus* transcriptome data
	R: 5'-ATATCTTCCCCAATTGGACTCCTC				
C0745	F: 5'-TAGGAAGTTCATCCATAAGCTCGG	$(TGA)_{10}$	NED	130-164 bp	Tranel lab *A. tuberculatus* transcriptome data
	R: 5'-CAATTCCAAGGAATCATCCTCATC				
C3561	F: 5'-CCATAAACCATTTTCCCAGACC	$(CCA)_8$	HEX	123-141 bp	Tranel lab *A. tuberculatus* transcriptome data
	R: 5'-ACTTCTGGCCCAATTAGGAAGTC				
C4999	F: 5'-CCACCCAATGACCCATACCTACTA	$(ACC)_8$	NED	120-141 bp	Tranel lab *A. tuberculatus* transcriptome data
	R: 5'-GATGAGGTTGATAATTGGGGTTCA				
AAC1	F: 5'-CCCACCAAGGATGATCATTTAGAC	AAC	6FAM	112-130 bp	Lee et al. 2009
	R: 5'-TCATCATTATTTGTTGGCGTTGAC				
TAG5	F: 5'-GTCGCTGAATTGTTTTAGCTTGGT	TAG	HEX	132-163 bp	Lee et al. 2009
	R: 5'-TGGGAATTCTCTCTTGTGACACAGT				
ATC9	F: 5'-TAGCCATTTCAACCTTACGAGGAA	ATC	NED	142-160 bp	Lee et al. 2009
	R: 5'-ACCGTTGATTGATTTTATGGCATC				
C3695	F: 5'-TCAACTTCTTATTCTTGGGTTGCTTC	$(TGA)_8$	6FAM	127-174 bp	Tranel lab *A. tuberculatus* transcriptome data
	R: 5'-CCTTACCTTCTCTCAAAAGCACCA				
C9333	F: 5'-AACTAAACGCATTTGCCATTGAA	$(GAT)_8$	HEX	165-199 bp	Tranel lab *A. tuberculatus* transcriptome data
	R: 5'-TGTTCATCTAACCACATCATAATGGAA				

Table S3. Population genetic statistics for each locus summarized over all populations. Na = number of alleles, Ne = effective number of alleles (estimated reciprocal of homozygosity), Ho = observed heterozygosity, He = expected heterozygosity.

Locus	Sample size	Na	Ne	Ho	He
C1140	722	20	8.86	0.82	0.89
C4097	720	6	2.04	0.49	0.51
C0745	702	14	5.87	0.76	0.83
C3561	722	7	1.55	0.35	0.35
C4999	716	9	3.02	0.58	0.67
AAC1	714	6	2.46	0.43	0.59
TAG5	710	10	2.55	0.46	0.61
ATC9	716	6	1.09	0.09	0.09
C3695	718	16	9.10	0.85	0.89
C9333	704	13	6.46	0.73	0.85

Online Supplemental

Click here to download Online Supplemental: Supplementary Figures.pptx

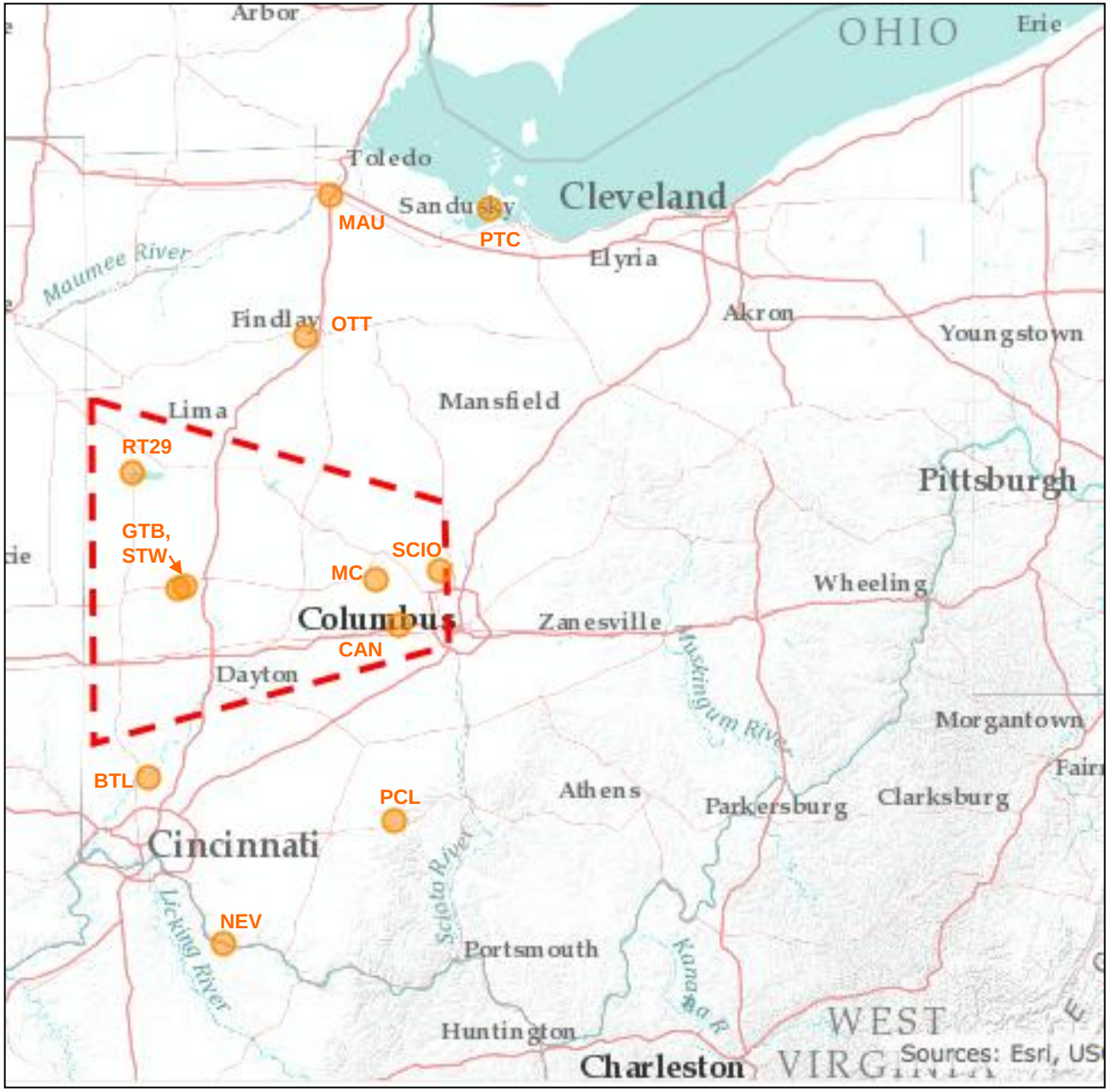
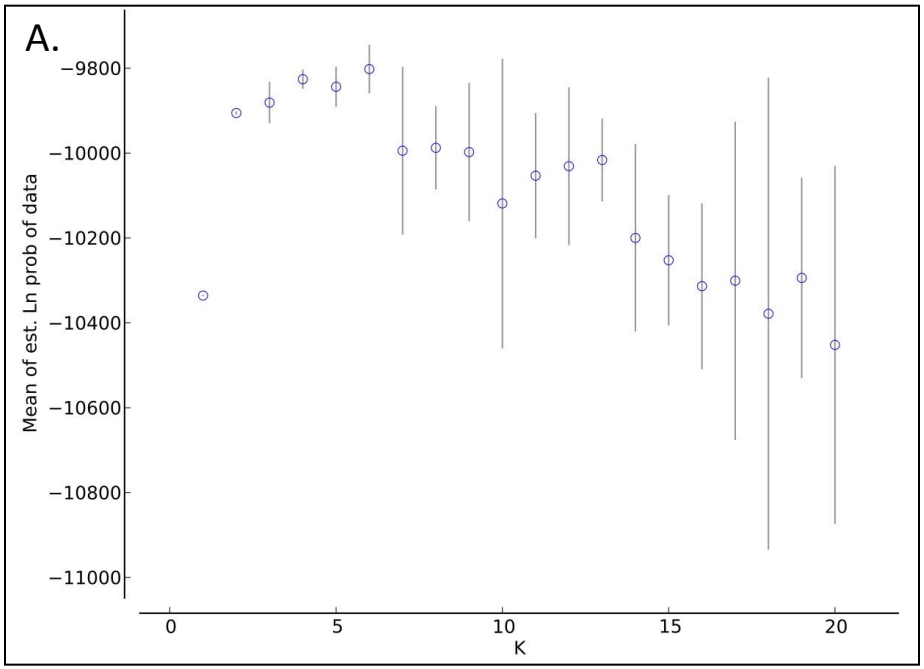


Figure S1.


A.
Mean of est. Ln prob of data
−9800
−10000
−10200
−10400
−10600
−10800
−11000
0
5
10
15
20
K

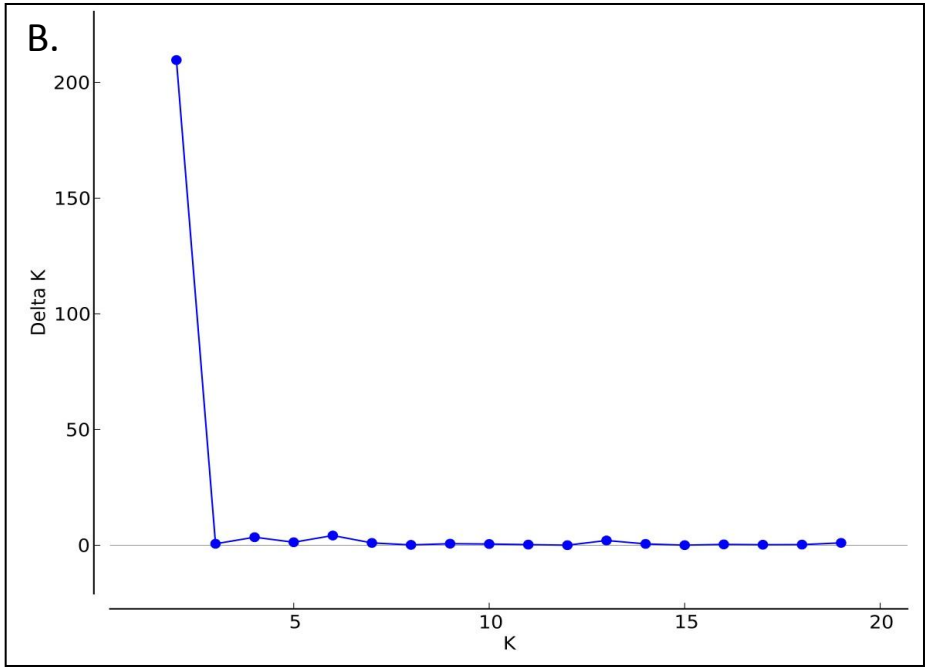
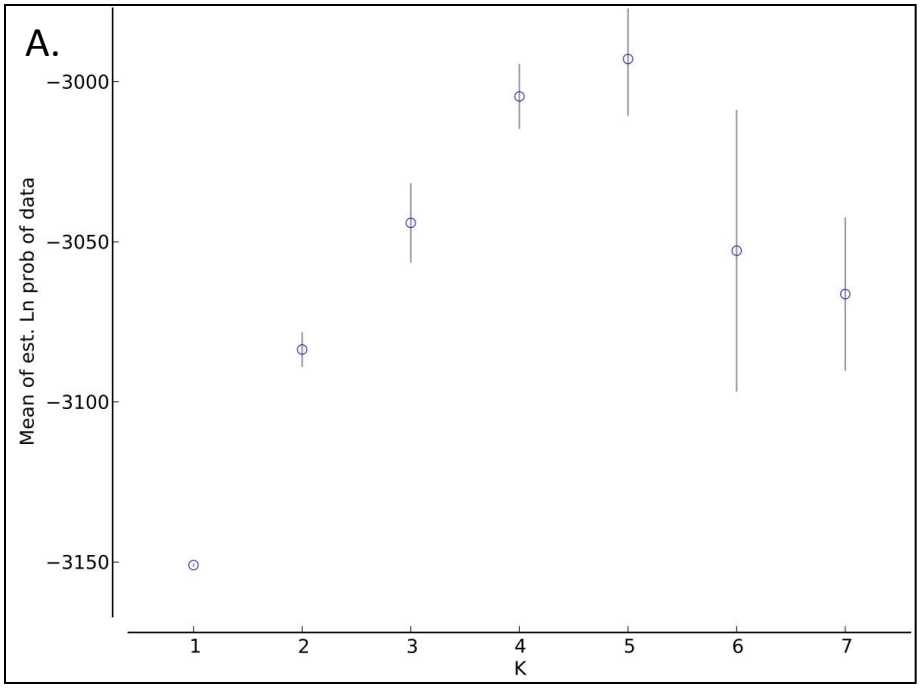

B.
Delta K
200
150
100
50
0
5
10
15
20
K

Figure S2.


A.
−3000
−3050
−3100
−3150
Mean of est. Ln prob of data
1
2
3
4
5
6
7
K

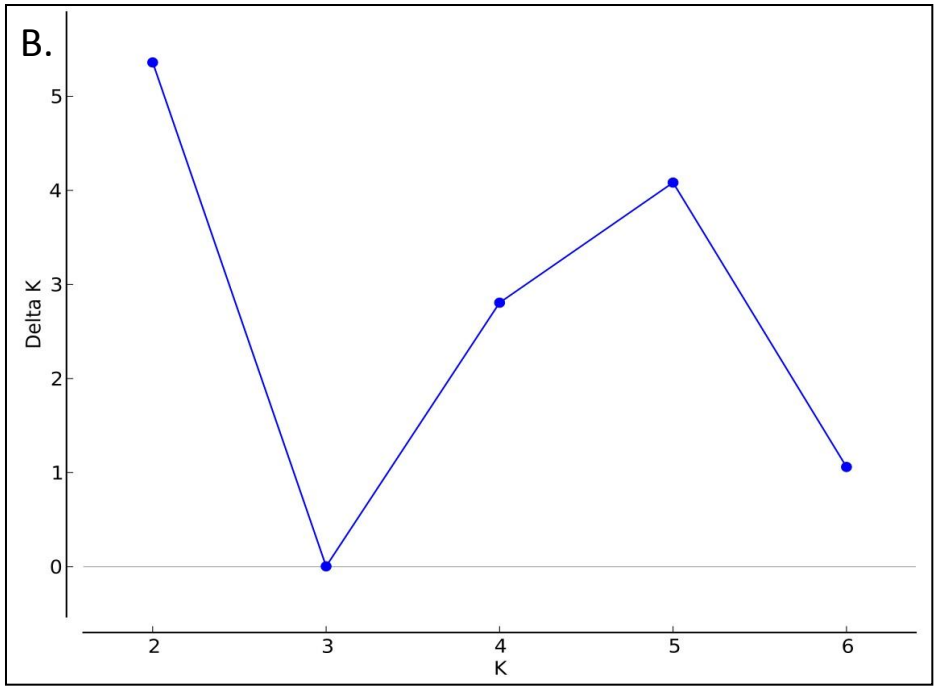

B.
5
4
3
2
1
0
Delta K
2
3
4
5
6
K

Figure S3.

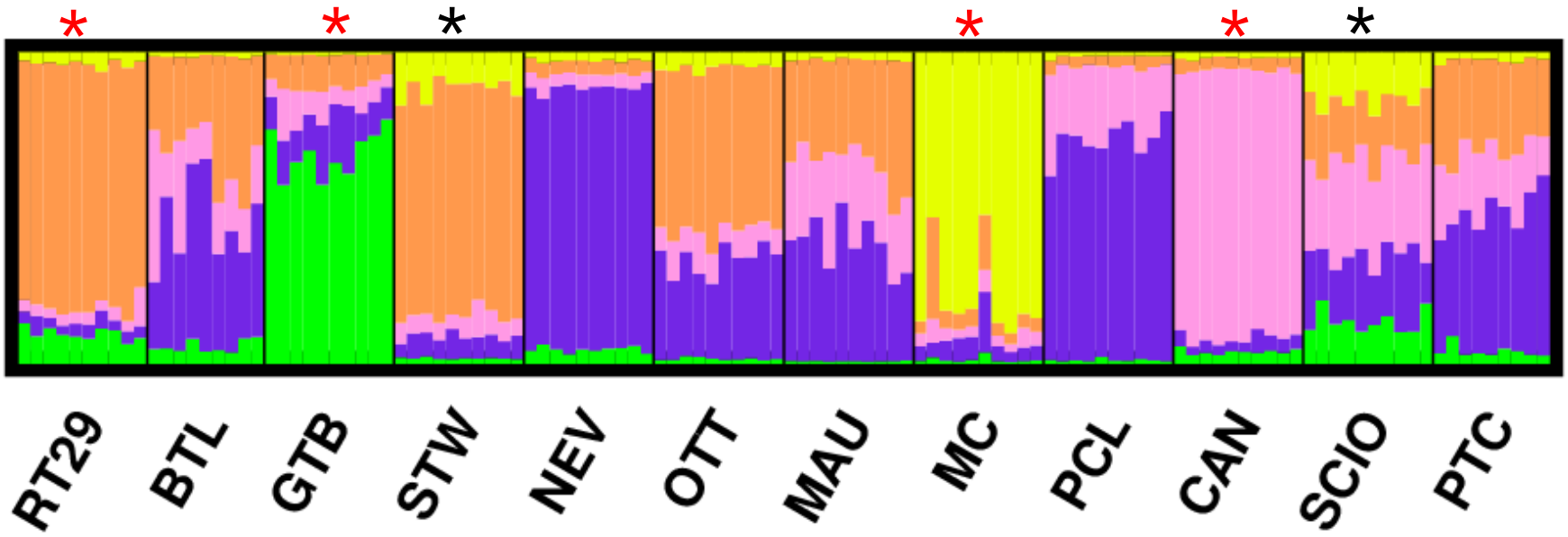


Figure S4.