Title
Tracing the Geographic Origins of Weedy Ipomoea purpurea in the Southeastern United States

Journal
Journal of Heredity, 104(5)

ISSN
0022-1503

Authors
Fang, Zhou
Gonzales, Ana M
Durbin, Mary L
et al.

Publication Date
2013-09-01

DOI
10.1093/jhered/est046

Peer reviewed

Journal of Heredity Advance Access published July 26, 2013

Journal of Heredity
doi:10.1093/jhered/est046

Tracing the Geographic Origins of Weedy *Ipomoea purpurea* in the Southeastern United States

Zhou Fang, Ana M. Gonzales, Mary L. Durbin, Kapua K. T. Meyer, Beau H. Miller, Kevin M. Volz, Michael T. Clegg, and Peter L. Morrell

From the Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108 (Fang, Gonzales, Miller, Volz, and Morrell); and the Department of Ecology and Evolutionary Biology, University of California, Irvine, CA (Durbin, Meyer, and Clegg).

Address correspondence to Peter L. Morrell, Department of Agronomy and Plant Genetics, University of Minnesota, 1991 Upper Buford Circle, 411 Borlaug Hall, St. Paul, MN 55108, or e-mail: pmorrell@umn.edu.

Abstract

Ipomoea purpurea (common morning glory) is an annual vine native to Mexico that is well known for its large, showy flowers. Humans have spread morning glories worldwide, owing to the horticultural appeal of morning glory flowers. *Ipomoea purpurea* is an opportunistic colonizer of disturbed habitats including roadside and agricultural settings, and it is now regarded as a noxious weed in the Southeastern US. Naturalized populations in the Southeastern United States are highly polymorphic for a number of flower color morphs, unlike native Mexican populations that are typically monomorphic for the purple color morph. Although *I. purpurea* was introduced into the United States from Mexico, little is known about the specific geographic origins of US populations relative to the Mexican source. We use resequencing data from 11 loci and 30 *I. purpurea* accessions collected from the native range of the species in Central and Southern Mexico and 8 accessions from the Southeastern United States to infer likely geographic origins in Mexico. Based on genetic assignment analysis, haplotype composition, and the degree of shared polymorphism, *I. purpurea* samples from the Southeastern United States are genetically most similar to samples from the Valley of Mexico and Veracruz State. This supports earlier speculation that *I. purpurea* in the Southeastern United States was likely to have been introduced by European colonists from sources in Central Mexico.

Key words: *geographic origin, nucleotide diversity, population structure, weedy species*

Ipomoea purpurea is native to the subtropics of Mesoamerica (Mabberley 1997) and has been spread worldwide by humans as an ornamental plant. Several New World species of morning glory, including *Ipomoea nil*, *I. purpurea*, and *Ipomoea tricolor*, have successfully adapted to tropical and warm temperate regions in Asia, Australia, and Europe, as well as North and South America, due to high phenotypic plasticity and genetic adaptability (Auld and Medd 1987; Defelice 2001). Naturalized populations are most abundant in the Southeastern United States (hereafter SE US) (Defelice 2001). *Ipomoea purpurea* occurs very infrequently outside of horticultural settings in drier regions such as the South Central and Southwestern US. *Ipomoea purpurea* and *Ipomoea hederacea* occur sympatrically as weeds in the SE US, but seed production is reduced when the species hybridize (Ennos 1981; Stucky 1985).

Ipomoea purpurea is a flowering vine that is cultivated as an ornamental because of the appeal of its large, showy flowers (Defelice 2001). In the SE US, the color of *I. purpurea* flowers ranges from completely white to pink and red and on to dark purple, whereas in Mexican natural populations flower color is typically purple (Brown and Clegg 1984), and populations associated with human habitation and agriculture are polymorphic (Ennos and Clegg 1983; Clegg and Durbin 2000). *Ipomoea purpurea* also possesses a number of traits that adapt the species to dispersal and a weedy habit. These include rapid growth, high seed output, and extended seed dormancy in soil seed banks (Baucom et al. 2011). Owing to an aggressive growth habit as a climbing vine, morning glory overgrows field corn, soybean, and other economically valuable plants and thus is considered a noxious weed (Webster and Coble 1997; Baucom and Mauricio 2008).

Populations of *I. purpurea* in the SE US were likely introduced from the native range of the species in Central and Southern Mexico (Mabberley 1997; Halvorson and Guertin 2003), but the timing and specific origins of introduction to the SE US remain a subject of speculation (Clegg and Durbin

2000). One possibility is that *I. purpurea* may have been introduced into the United States along with maize cultivation ~4000 years ago and then spread to the SE US (Hill 2001; Merrill et al. 2009). A second scenario is that the species was introduced into the SE US through Europe, after European contact with the New World. Under this hypothesis, *I. purpurea* seeds are believed to have been collected from native fields by early Spanish colonizers and sent back to Spain where they were planted in monastery gardens (Defelice 2001; Halvorson and Guertin 2003). *Ipomoea purpurea* evidently arrived in England by 1621 (Halvorson and Guertin 2003). Around 1700, it was brought to the American colonies as a horticultural introduction and spread by way of garden peddlers. It is postulated to have quickly escaped cultivation to become a naturalized weed (Halvorson and Guertin 2003). Therefore, *I. purpurea* may have been introduced into the SE US from the native populations in Mesoamerica by way of European colonists in the last several hundred years, a much more recent introduction than expected if it was introduced as a commensal of maize cultivation.

Haplotype data from DNA resequencing have the potential to provide insight into the origins of weedy morning glory populations in the SE US. These high-resolution data provide a powerful means to detect population structure (Morrell and Clegg 2007; Lawson et al. 2012). There are several questions that can be illuminated from haplotype data that are key in understanding the scenario for the introduction of *I. purpurea* in the SE US and the geographic origins of the SE US population. First, what is the extent of genetic differentiation among Mexican populations? Second, was genetic diversity in the SE US populations reduced by an introduction bottleneck? Third, can the SE US haplotypes be traced back to specific geographic regions of Mexico? To address these questions, we analyze resequencing data from 11 loci based on 30 accessions sampled from the native range of the species in Mexico (Gonzales et al. 2012) and 8 *I. purpurea* samples from the SE US.

Materials and Methods

Materials

Our sample of 38 *I. purpurea* accessions includes 30 single individuals representing populations in the native range of the species in Mexico (Villasenor and Espinosa 1998; Gonzales et al. 2012) and 8 representative samples from the SE US (see Supplementary Table S1 online). Seeds were collected from a wide geographic range of populations in Mexico over a number of years ranging from 1986 through 2002. Individual seed accessions were periodically rejuvenated in a greenhouse in California to maintain the viability of each accession. Earlier studies have shown that there is relatively little differentiation among southeastern populations of *I. purpurea* (Epperson and Clegg 1986; Huttley et al. 1997). The 8 samples from the SE US include 6 accessions from various flower color variants, which were collected from distinct populations spanning ~1000 km^2 of Northern Georgia (Epperson and Clegg 1986) and 2 accessions collected from naturalized populations near Athens and Lumberton, NC. Although sampling does affect the potential for inference of demographic history among SE US samples and is likely too small for accurate estimation of recombinational diversity (Morrell et al. 2006), the sample size of 8 individuals in the SE US is appropriate for estimation of nucleotide sequence diversity (Pluzhnikov and Donnelly 1996; Felsenstein 2006) and the effect of introduction on diversity at multiple loci. We have also resequenced accessions of *Ipomoea alba* and *I. nil,* closely related sister species of *I. purpurea* (Miller et al. 1999), to infer the ancestral state of mutations at each locus. The 11 resequenced loci are the same as those reported by Gonzales et al. (2012): anthocyanidin synthase (*ALS*), 2 chalcone synthases (*CHS-D*, *CHS-E*), flavanone-3-hydroxylase (*F3H*), flavonol synthase (*FLS*), dihydroflavonol 4-reductase B (*DFR-B*), bHLH transcriptional regulator (*IpbHLH1*), 2 R2R3 Myb transcriptional regulators (*IpMyb1* and *IpMyb4*), and *WD40* repeat protein (*IpWDR1*), and UDP-glucose flavonol 3-0-glucosyl transferase (*UF3GT*). Nine of the loci are known to participate in a pathway that produces floral pigments and were identified in studies focused on the genetic basis of floral pigment variation (Durbin et al. 2000; Lu and Rausher 2003; Toleno et al. 2010).

Sanger sequencing was used to obtain high-quality consensus sequence from PCR products for each accession. Sequence reads were assembled using phred and phrap with crossmatch used to screen out cloning vector (Ewing and Green 1998; Ewing et al. 1998). Consed (Gordon et al. 1998) was used for visualizing the reads. PolyPhred (Bhangale et al. 2006) was used to detect single-nucleotide polymorphisms (SNPs). When a sample produced a heterozygous amplicon, the 2 haplotypes were phased experimentally (Chen et al. 2010) using cloning with Qiagen pDrive Cloning Vector. At least 3 clones were sequenced per haplotype. The accuracy of inferred haplotypes was tested using error detection on triplets of nucleotide sites in the program EDUT (Toleno et al., 2007).

Population Structure

We used the program STRUCTURE (Pritchard et al. 2000) for genetic assignment within the sample of *I. purpurea* accessions. STRUCTURE uses a Bayesian-model-based clustering method to infer population structure using genotype (infinite allele) data. STRUCTURE assumes there are *K* clusters for the samples and assigns the individuals to clusters based on distinct allele frequencies. We treated each of these 11 loci as a series of quasi-independent chromosomal segments as defined by direct evidence of recombination based on the 4-gamete test (Hudson and Kaplan 1985), similar to the approach reported by Morrell and Clegg (2007) and Chen et al. (2009). Under the infinite sites model (Kimura and Crow 1964), there are 4 possible gamete types (or haplotypes) when there are 2 alleles at each of 2 sites. If all 4 combinations are present, the 4-gamete test indicates recombination among sites (Hudson and Kaplan 1985) because more than 1 mutation at a single-nucleotide site is assumed to have a probability approaching 0. We use the 4-gamete test as a conservative test indicating

recombinational independence between chromosomal segments. Each of the chromosomal segments contains all the SNPs within a 4-gamete interval with unique haplotypes identified as alleles while excluding singleton SNPs, because SNPs observed at a single locality are not geographically informative. For STRUCTURE analysis, we explored models using both an admixture and no admixture model and with both correlated and uncorrelated allele frequencies with $K = 1$–5 clusters. Because the no admixture and uncorrelated allele frequency models resulted in higher likelihoods, they were used for the final analyses. For each value of K, we used 10 replicate runs, with both burn-in and run lengths of 100 000 iterations. The program CLUMPP (Jakobsson and Rosenberg 2007) was used to summarize assignment results across replicate runs.

In an initial round of analysis, STRUCTURE was used only for the 30 Mexican accessions. The best K value was estimated using the ad hoc value ΔK, which is based on the second-order rate of change of the likelihood function with respect to K (Evanno et al. 2005). To identify likely origins of the SE US samples, we then included all 38 accessions in genetic assignment analysis with Mexican accessions treated as a learning sample. Individuals MTC7 and MTC241 were excluded from the learning sample as they have primary assignment outside the region in which they were sampled.

The program Infocalc was used to calculate the informativeness for assignment for each haplotype segment used for genetic assignment. Informativeness for ancestry coefficients (I_a), informativeness for assignment (I_n), and optimal rate of correct assignment from both 1-allele and 2-allele estimates (Rosenberg et al. 2003; Rosenberg 2005) were computed using the clusters identified by STRUCTURE. Among these 4 values, I_n is most directly relevant here. This statistic is based on the degree to which each locus (or haplotype segment) contributes to distinction among populations. A haplotype with a higher value of I_n indicates that the haplotype is more informative for genetic assignment.

The nearest neighbor statistic (S_{nn}) (Hudson 2000) and a statistical test for detecting differentiation in subpopulations (F_{ST}) (Hudson et al. 1992) as implemented in libsequence (Thornton 2003) were used to test for evidence of geographic structure at individual loci.

To determine the degree of differentiation among populations, we also examined the number of shared, fixed, and private polymorphisms in each of the Mexican populations and SE US accessions. For our data, this analysis could be conducted using individual SNPs, haplotype segments (which have the advantage of partial recombinational independence), and full-length haplotypes from resequencing. Sampled haplotypes, SNP variation contributing to each haplotype, and the geographic origins of each haplotype were visualized using SNAP MAP (Aylor et al. 2006).

Sequence Diversity Analysis

Levels of nucleotide sequence diversity were examined using estimates of $\theta = 4 N_e\mu$, including the number of segregating sites (θw) (Watterson 1975) and the proportion of pairwise difference per locus ($\theta\pi$) (Tajima 1983), where N_e is the effective population size, and μ is the mutation rate per generation. We report tests of neutrality, including Tajima's D (the normalized difference between $\theta\pi$ and θw) (Tajima 1989) and Fay and Wu's H (Fay and Wu 2000). Fay and Wu's H incorporates information on the derived versus ancestral state of mutations, so an accession of *I. alba* or *I. nil* was used as an out-group for each locus. The impact of recombination and extent of linkage disequilibrium were estimated using the 4-gamete test (reported as R_m) (Hudson and Kaplan 1985). Sequence summary statistics were calculated using tools from the libsequence C++ library (Thornton 2003).

A multilocus maximum-likelihood estimate of θw was calculated across all loci in a sample using the recursion equations in Hudson (1991). This method assumes a constant mutation rate across loci and makes the (conservative) assumption of no intralocus recombination. This estimator is implemented in the program Theta Curve (Ross-Ibarra et al. 2009).

Results

Genetic Assignment Analysis

Results of the nearest neighbor test for geographic structure S_{nn} and F_{ST} indicate that the combined Mexican sample and the SE US sample are genetically differentiated (Table 1). For 9 of 11 loci, S_{nn} is significant at $P = 0.05$. F_{ST} is significant for all 11 loci. This establishes that the SE US samples are not a representative sample of the total Mexican gene pool.

We next ask whether the Mexican accessions exhibit geographic substructure as a prelude to asking where in Mexico the SE US accessions might have originated. To test for population structure among the Mexican populations, we used genetic assignment in STRUCTURE and found significantly higher ΔK values for $K = 2$ and 3 clusters than for other K values (see Supplementary Figure S1 online). For $K = 2$, the Mexican samples are divided into the geographic region running from Xapala in Veracruz State (east central Mexico) to the Valley of Mexico (except for accessions MTC7, MTC275

Table 1 Results of the nearest neighbor test for geographic structure, S_{nn}, and test of allele frequency differentiation, F_{ST}

	S_{nn}	*P*-value	F_{ST}	*P*-value
ALS	0.875	0.000	0.247	0.000
CHS-D	0.812	0.007	0.051	0.024
CHS-E	0.919	0.000	0.288	0.000
DFR-B	0.950	0.000	0.187	0.000
F3H	0.921	0.000	0.060	0.026
FLS	1.000	0.000	0.227	0.000
IpbHLH1	0.695	0.236	0.075	0.041
IpMyb1	0.845	0.001	0.185	0.000
IpMyb4	0.745	0.068	0.183	0.002
IpWDR1	0.903	0.000	0.153	0.000
UF3GT	0.953	0.000	0.148	0.000

P-values are based on 1000 permutations. The samples are divided into accessions from SE US versus Mexico.

from Veracruz and the Valley of Mexico). The second division includes accessions from central and southern Mexico. For $K = 3$, samples cluster into geographic groups, with an Eastern (eastern portion of Distrito Federal near Mexico City, and Veracruz State), Southern (all samples from Chiapas), and Western Mexican groups (the western portion of the Valley of Mexico and the majority of samples from Morelos and Oaxaca; Figure 1). There are 2 exceptions. MTC7 in the Eastern group and MTC241 in the Western group show >90% probability of assignments to the Southern group (Figure 1).

Consistent with an earlier study by Huttley et al. (1997), the Southern population from Chiapas has the lowest level of nucleotide sequence diversity, lending further support to the identification of the Southern cluster as a separate population. The identification of 3 relatively distinct populations may provide greater precision to infer geographic origins. Moreover, the S_{nn} test among Mexican populations indicate differentiation at all loci for the Eastern and Southern populations, 10 of 11 loci for the Western and Southern populations, 8 of 11 loci for the Eastern and Western populations (see Supplementary Table S2 online) and each of the 3 Mexican populations is also genetically highly differentiated from the SE US accessions (see Supplementary Table S3 online). Therefore, the 3 clusters, the Eastern, Southern, and Western Mexican, are treated as distinct populations for further analyses.

Given that the pooled Mexican samples and the SE US samples are differentiated (Table 1) and that geographic structure is evident in the Mexican samples (Figure 1), we used the clusters inferred from the Mexican accessions to assign the SE US accessions to regions of origin in Mexico. The SE US accessions share more similarity to the Mexican Western population than the Eastern and Southern populations (Figure 1).

To evaluate the strength of the assignment results, we calculated the informativeness for assignment (I_n) for all 89 haplotype segments to the 4 populations (see Supplementary Table S4 online). The average I_n for all 89 haplotype segments is 0.37. Two segments of the *DFR-B* locus have $I_n = 1$, indicating they are completely informative for assignment.

Nucleotide Sequence Diversity

We now consider nucleotide sequence diversity levels in the SE US and in Mexico. Among the 8 samples in the SE US, there are no polymorphic sites at *CHS-E*, *FLS* and *IpMyb4* (Table 2). Diversity based on θw and $\theta\pi$ is much higher at *DFR-B* and *UF3GT* than at other loci (Table 2).

The SE US samples have dramatically lower nucleotide diversity compared with samples from Mexico (Table 2). Pairwise diversity ($\theta\pi$) is much lower for 10 of 11 loci. The exception is *UF3GT*, which has 2 haplotypes at equal frequency in the SE US samples, thus diversity is slightly higher for this locus in the SE US samples compared with other loci (Figure 2). The haplotype diversity for all 11 loci is also much lower for the SE US samples than in Mexican samples (see Supplementary Figure S2 online). A maximum-likelihood estimate of θw indicates a significant reduction of diversity in the SE US relative to Mexican populations (Figure 3). θw is much lower for the SE US samples compared with all the samples in Mexico and each of the 3 Mexican populations.

Origin of the SE US Population

The proportion of shared SNPs (S_S) between the SE US population and the Mexican Eastern and Western populations is larger than between the SE US population and the Mexican Southern population (Figure 4). Average F_{ST} between the SE US population and the Mexican Eastern and Western populations (0.22 and 0.15, respectively) is much lower than that between the SE US population and Mexican Southern population (0.32), which also reveals that the SE US population is more similar to the Mexican Eastern and Western clusters.

It is informative to consider the pattern of shared haplotype segments between the SE US and the Mexican samples. Haplotypes include multiple allelic states, and thus, some alleles can be unique to geographic regions. Thus, based on simulation and empirical studies they are more informative than SNPs (Gattepaille and Jakobsson 2012). MTC275 from Veracruz State has the highest proportion of haplotype sharing (0.68) with SE US samples followed by MTC196 (0.58) from Morelos and MTC4 (0.57) from Mexico City (see Supplementary Figure S3 online).

Across all loci, 11 of 22 full-length haplotypes observed in the SE US accessions are also found in the Mexican Eastern population, whereas only 6 haplotypes are found in the Mexican Western population. Among haplotypes that differ between the SE US and Mexican samples by 1 or 2 mutations and haplotypes that share identical mutation patterns between the SE US and Mexican samples, 11 are similar to the Western population and only 4 haplotypes are similar to the Eastern population (Figure 5; see Supplementary Figure S4 online). The relatively even representation of these 2 Mexican populations (with 15 haplotypes from Eastern, 17 from Western) indicates that genetic diversity in the SE US was probably contributed from both regions, but the larger proportion of identical full-length haplotypes suggests that the contribution from the Eastern population was more recent. For *DFR-B*, the samples from the SE US have 2 haplotypes. An accession from the Western cluster of Mexico (MTC285) shares the same haplotype as 3 samples from the SE US (Figure 5A). The nucleotide diversity of this locus is significantly higher than other loci (Table 2), but we find the same haplotype between the SE US and the western cluster, which gives strong evidence that the SE US population and western cluster of Mexico are genetically similar to each other.

Discussion

The questions addressed in this investigation are 2-fold: first, which of the 2 scenarios for the introduction of *I. purpurea* into the SE US is best supported by haplotype data? Second, can anything be said about the geographic regions

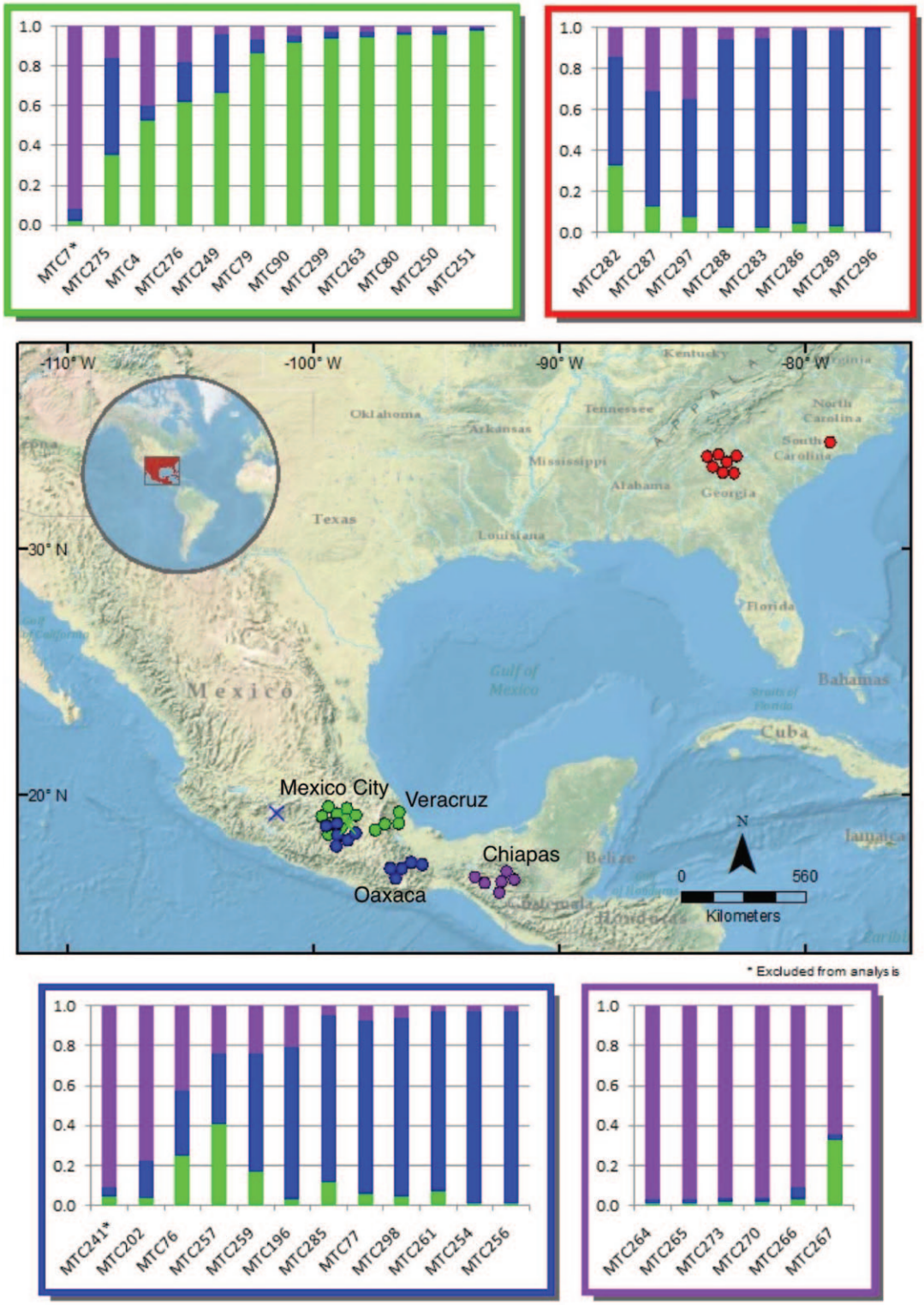

Figure 1. The geographic distribution of sampled *I. purpurea* accessions. Genetic assignment analyses indicate 3 Mexican populations, Eastern, Southern and Western, shown in green, purple, and blue, respectively. The inclusion of samples within the colored groups was determined primarily by genetic assignment analysis and also accounted for geographic proximity. The genetic assignment barplots correspond to each of these regions. The SE US population is in the red box. The barplot shows the probability of assignments of each sample into the 3 Mexican populations as indicated by the *y*-axis. MTC7 and MTC241 (the crosses on the map) were excluded from the learning sample for the assignment of the SE US accessions.

Table 2 Summary of measures of nucleotide diversity at all 11 loci

	P	*N*	*L*	*S*	*h*	*H*	θw	$\theta\pi$	Tajima's *D*	Fay/Wu's *H*	R_m
ALS	US	9	808	3	3	0.42	0.001	0.001	−0.94	−0.005	0
	E	12	808	9	5	0.67	0.004	0.004	0.12	−0.001	1
	S	7	808	7	4	0.81	0.004	0.004	0.52	−0.002	0
	W	16	808	18	13	0.98	0.007	0.005	−1.07	0.004	1
CHS-D	US	8	903	4	2	0.54	0.002	0.002	1.70	0.002	0
	E	13	903	16	9	0.91	0.006	0.006	0.10	−0.001	1
	S	6	903	11	4	0.87	0.005	0.006	0.81	0.002	0
	W	14	903	21	11	0.96	0.007	0.005	−1.20	0.001	2
CHS-E	US	8	894	0	1	0	0	0	NA	0	NA
	E	13	894	25	8	0.86	0.009	0.008	−0.58	−0.009	3
	S	8	894	10	2	0.54	0.004	0.006	1.93	−0.001	0
	W	15	894	35	13	0.98	0.012	0.013	0.17	−0.008	9
DFR-B	US	8	1020	30	2	0.54	0.012	0.017	2.06	0.001	0
	E	12	1247	46	5	0.73	0.018	0.019	0.23	−0.006	5
	S	7	1247	29	3	0.76	0.013	0.019	2.23	0.003	0
	W	13	1247	66	10	0.96	0.027	0.023	−0.71	0	8
F3H	US	9	954	17	4	0.58	0.007	0.006	−0.11	−0.001	2
	E	12	961	19	7	0.77	0.007	0.006	−0.34	−0.003	3
	S	6	961	1	2	0.33	0.000	0.000	−0.93	0	NA
	W	14	961	24	12	0.98	0.008	0.008	0.23	0.005	6
FLS	US	8	1091	0	1	0	0	0	NA	0	NA
	E	12	1113	17	7	0.89	0.005	0.006	0.38	0.003	2
	S	6	1113	1	2	0.33	0.000	0.000	−0.93	0	NA
	W	13	1113	28	10	0.95	0.008	0.009	0.27	0.003	6
IpbHLH1	US	8	1230	3	2	0.54	0.001	0.001	1.60	0	0
	E	12	1230	7	3	0.53	0.002	0.002	−0.33	0	0
	S	6	1230	8	5	0.93	0.003	0.003	−0.06	0.001	0
	W	14	1230	13	6	0.68	0.003	0.002	−1.04	−0.001	1
IpMyb1	US	8	848	4	2	0.25	0.002	0.001	−1.54	−0.005	0
	E	13	1143	36	9	0.91	0.017	0.021	1.11	0.002	6
	S	6	1143	30	3	0.60	0.016	0.016	−0.16	−0.001	0
	W	12	1143	42	8	0.89	0.019	0.016	−0.88	−0.003	5
IpMyb4	US	8	468	0	1	0	0	0	NA	0	NA
	E	14	489	19	6	0.80	0.014	0.018	1.49	0	0
	S	6	489	24	4	0.80	0.025	0.022	−0.66	−0.005	0
	W	12	489	19	7	0.88	0.014	0.014	−0.05	−0.010	0
IpWDR1	US	8	1145	4	2	0.25	0.001	0.001	−1.54	−0.002	0
	E	14	1145	21	10	0.95	0.006	0.005	−0.82	−0.004	3
	S	6	1145	8	3	0.60	0.003	0.003	−0.74	−0.001	0
	W	16	1145	21	13	0.98	0.006	0.005	−0.30	−0.004	4
UF3GT	US	8	1081	19	2	0.57	0.007	0.011	2.50	0.005	0
	E	13	1088	44	9	0.87	0.013	0.010	−1.03	0.003	4
	S	7	1088	18	3	0.67	0.007	0.007	−0.23	−0.001	0
	W	12	1088	38	11	0.98	0.012	0.009	−1.07	−0.003	2
Average	US	8	949	7.6	2	0.33	0.003	0.004	0.47	0	0.25
	E	13	1002	24	7	0.81	0.009	0.010	0.03	−0.001	2.6
	S	7	1002	13	3	0.66	0.007	0.008	0.16	0	0
	W	14	1002	30	10	0.93	0.011	0.010	−0.51	−0.001	4

h, the number of haplotypes; *H*, haplotype diversity; *L*, the length of the locus in bp; P, populations (SE US, E, Mexican Eastern population; S, Mexican Southern population; and W, Mexican Western population); *N*, sample size; NA, not applicable; *S*, the number of segregating sites; R_m, minimum recombination number.

within Mexico from which *I. purpurea* was introduced into the SE US? The 2 scenarios for the introduction of common morning glory into the SE US are 1) a gradual move northward over the last several thousand years along with the introduction of maize culture—termed the maize migration hypothesis, or 2) movement from Mexico to Europe and then to the SE US as a horticultural plant, following the European invasion of Mexico—termed the European migration hypothesis.

Scenario for the Introduction of *I. purpurea* into the SE US

Let us consider the maize migration hypothesis in greater detail. The peoples of Mesoamerica began to domesticate

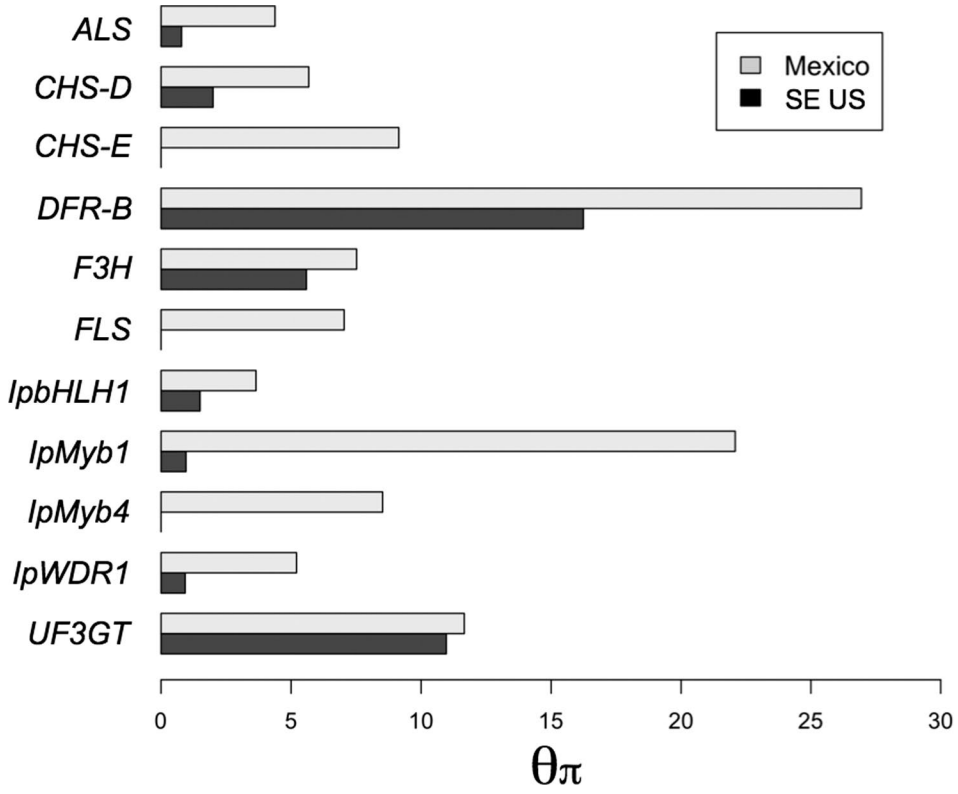

Figure 2. The comparison of estimated nucleotide sequence diversity $\theta\pi$ (per locus) between samples from Mexico (gray) and the SE US samples (black).

plants, including common bean, maize, and squash between 8000 and 10 000 years ago (Smith 2006; Piperno et al. 2009). Owing to its invasive climbing vine habit, *I. purpurea* is known, for as long as the historical records suffice, to have infested native maize milpas in many parts of Mexico. It is highly probable that the flower color diversity of common morning glory was selected by Neolithic cultivators of maize for the same reasons that diverse maize kernel mutations were selected—the esthetic appeal of color diversity (Clegg and Durbin 2003). Maize culture then expanded northward and eastward, driven by an increasing commitment to cultivation, eventually leading to the establishment of maize agriculture in the Southwestern United States by ~4000 years ago (Hill 2001) and into the Eastern United States (New York State) by as early as 2900 years ago (Hart et al. 2007). Owing to its weedy association with maize cultivation in Mexico, it is plausible that common morning glory migrated, as a semidomesticate of maize culture, along with the northward migration of maize cultivation. Thus, *I. purpurea* may have been brought to the United States along with domesticated maize. A question naturally follows: were the flower color mutations of *I. purpurea* selected in a single location or were the flower color morphs selected from recurrent mutations in different locations in Mexico by the Neolithic cultivators of maize? The latter scenario is plausible because many of the genes involved in flavonoid biosynthesis in *I. purpurea* harbor active transposable elements that cause high rates of mutation (Epperson and Clegg 1987; Clegg and Durbin 2000; Clegg and Durbin 2003).

Despite this, a number of lines of evidence appear to favor the European migration hypothesis. To begin, the geographic distribution of genetic diversity in *I. purpurea* appears paradoxical, featuring low levels of molecular (Figures 2 and 3; see Supplementary Figure S2 online) and biochemical polymorphism, but high levels of flower color diversity in the SE US relative to native Mexican populations (Glover et al. 1996; Clegg and Durbin 2000). Higher levels of flower color polymorphism in the SE US, compared with native Mexican populations where most populations are monomorphic for purple corolla flowers, is consistent with the hypothesis that the plant was introduced to the SE US based on the horticultural appeal of diverse flower colors (Epperson and Clegg 1986; Glover et al. 1996).

The data presented in this study quantify the magnitude of the reduction in genetic diversity in the SE US relative to Mexico and indicate a strong founder effect consistent with multiple founder events (e.g., Mexico to Europe and Europe to the SE US). The mean $\theta\pi$ at silent sites for the SE US samples (0.0071 per site) is 40.3% of $\theta\pi$ at silent sites for Mexican samples (0.0176 per site; Gonzales et al. 2012) and the maximum likelihood estimate of diversity across all loci indicate θw in the SE US is only 28.9% of that for the Mexican accessions (Figure 3). The reduced diversity in the SE US is consistent with previous allozyme data,

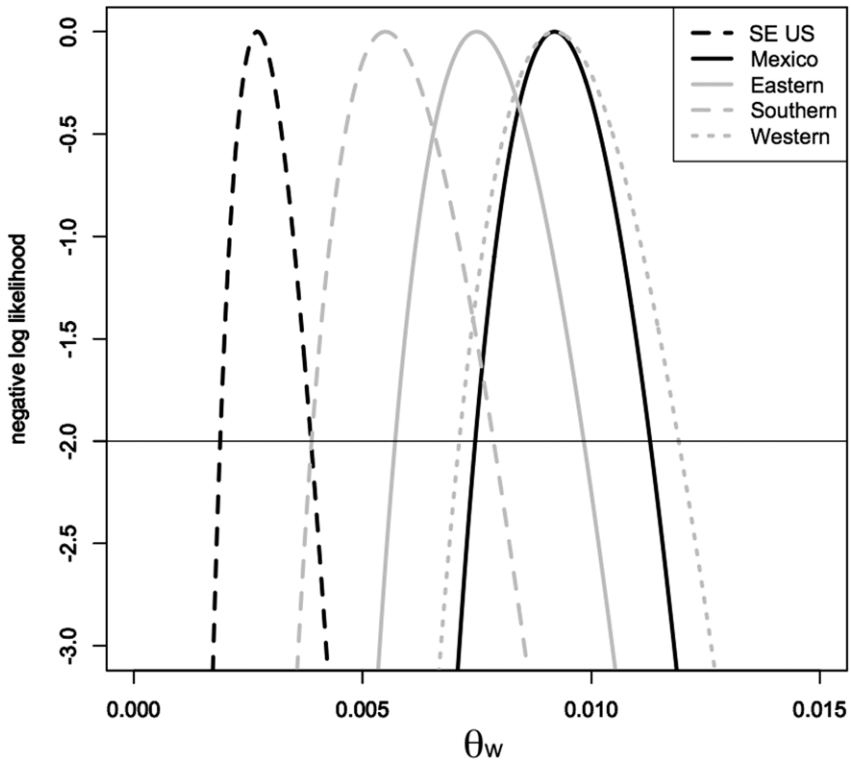

Figure 3. Maximum-likelihood estimates of nucleotide sequence diversity θw (per base pair) among the SE US and Mexican populations (black curves), as well as each of the 3 Mexican populations: Eastern, Southern, and Western (gray curves). The 95% confidence interval can be approximated by the intersection of each curve with the horizontal line at −2.

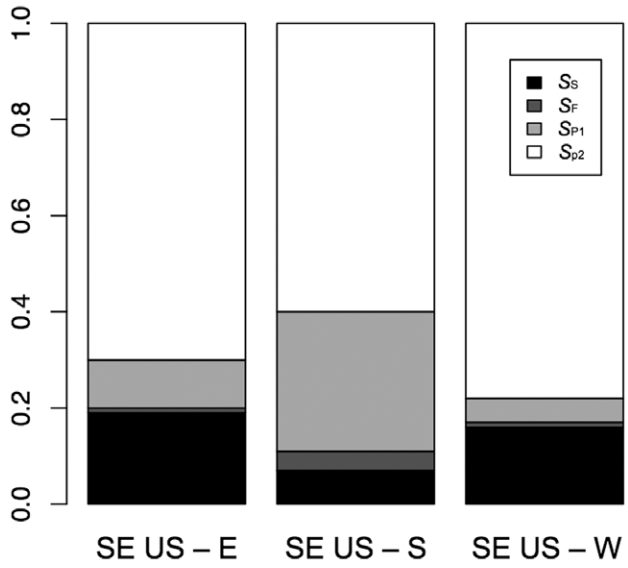

Figure 4. Comparison of proportions of shared (S_S), fixed (S_F), and unique private sites among each pairwise comparison between the SE US population and each of the 3 Mexican populations, Eastern (E), Southern (S), and Western (W). S_{P1}: the proportion of private sites in the SE US population; S_{P2}: the proportion of private sites in each of the 3 Mexican populations.

ribosomal DNA RFLP diversity, and nucleotide sequence diversity at *CHS-A* locus (Glover et al. 1996; Huttley et al. 1997). These studies also suggested that a severe population bottleneck accompanied the introduction of *I. purpurea* into the SE US. In contrast, maize diversity in the United States is high and does not suggest a strong founder effect (Buckler et al. 2006). Expectations for diversity in an introduced horticultural commensal species are unclear, but if maize and the common morning glory migrated together, it is difficult to account for the large difference in diversity given shared demographic history. In addition, the deficit of singleton and rare SNPs in the SE US samples are suggestive of a recent introduction of *I. purpurea* from Mexico into the SE US (Figure 5; see Supplementary Figure S4 online). Moreover, the high degree of similarity between samples from the SE US and Mexican Eastern population also supports the European migration hypothesis. For 9 out of 11 loci, we observe the sharing of full-length haplotypes between the SE US and Mexican Eastern populations. This would be less likely to occur if *I. purpurea* had been in the SE US for several thousand years as is the case for maize because mutation and intragenic recombination would have greater potential to alter introduced haplotypes. Finally, historical research in Defelice (2001) and Halvorson and Guertin (2003), reviewed in the Introduction, appear to support the European migration hypothesis.

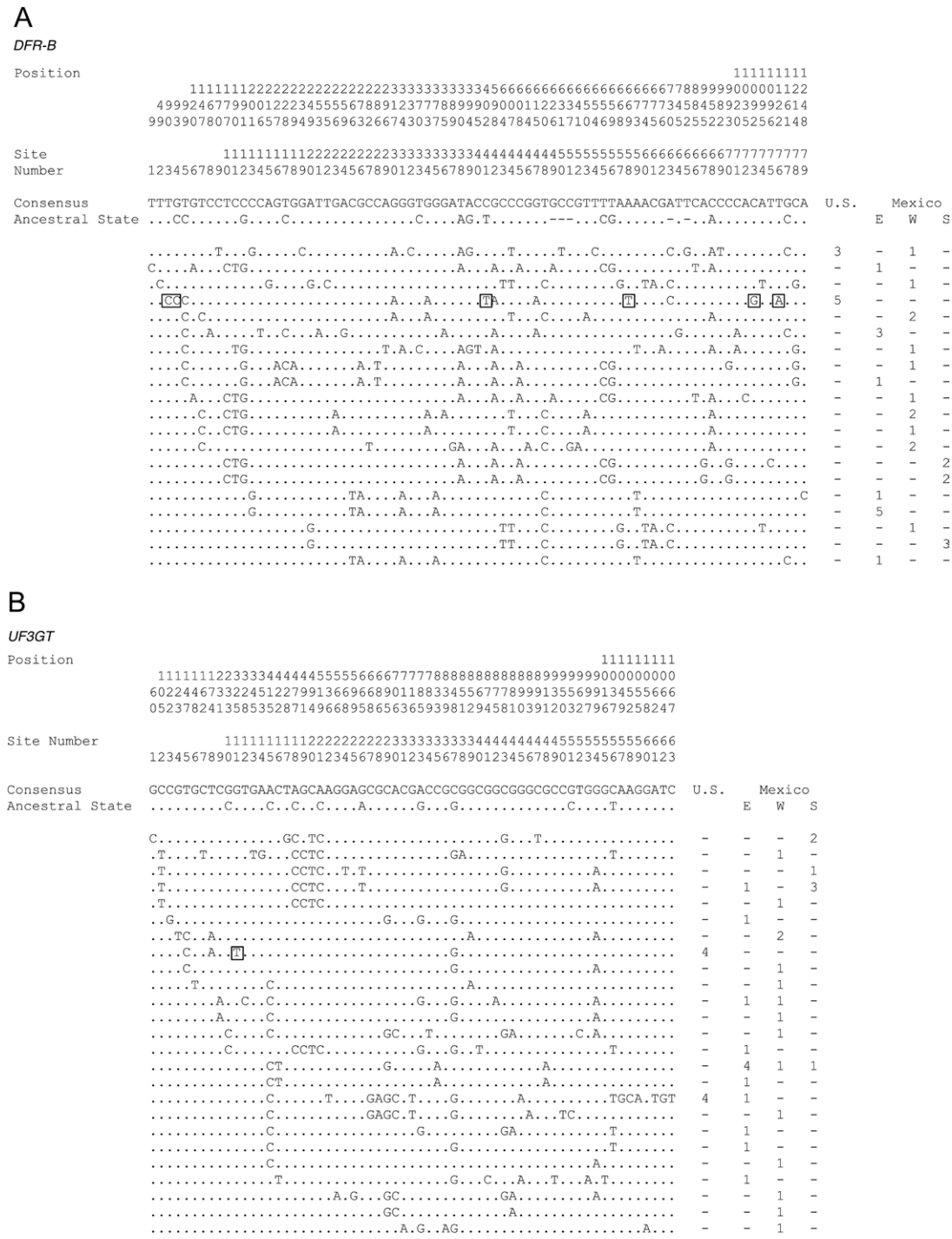

Figure 5. Haplotype alignments for 2 genes: (**A**) *DFR-B* and (**B**) UF3GT with SNPs private to one of the divergent haplotypes in the SE US population based on all unique haplotypes. Nucleotide positions for each segregating sites are shown in the first row. Nucleotide state is shown for difference from consensus with consensus shown as a point. Ancestral states are determined based on *I. alba* samples for these 2 loci. The haplotypes are sorted according to the positions of the appearance of the first nucleotide substitutions that differ from consensus. The nucleotide substitutions characterizing the haplotypes of the SE US and Mexican Eastern, Southern, as well as Western populations are shown on the right column. The number indicates how many times the haplotype appears in each of the 4 populations. A dash indicates the haplotype is absent in that population. Nucleotides highlighted by a box are private to the SE US population.

Geographic Origins of the SE US Accessions

We now consider whether the geographic origins of the SE US populations can be inferred from haplotype data. The haplotype data suggest that the SE US populations originated along the axis from the Valley of Mexico (the intersection of Eastern and Western populations) to Veracruz State at about 19.5° latitude (and more specifically from the region of Xapala in Veracruz State). This inference is supported by samples (MTC275) collected at the outskirts of Xapala and (MTC4) from Mexico City and (MTC196) from Morelos that show the closest relationship to the SE US samples based on the proportion of shared haplotype segments (see Supplementary Figure S3 online). Moreover, a unique haplotype at *UF3GT* is present in both the SE US population and a sample from Mexico City (the intersection of Eastern and Western populations; Figure 5B). Huttley et al. (1997) also found a unique SE US allele in a chalcone synthase gene present in a sample from Mexico City. The locus *DFR-B*, with the highest nucleotide diversity has a unique haplotype present in both the SE US and MTC285 from Morelos (south of Mexico City; Figure 5A).

Although the haplotype data suggest that both Mexican Eastern and Western populations are likely to have contributed to the SE US accessions, genetic assignment and F_{ST} analysis indicate the SE US accessions are more similar to the Mexican Western population than the Eastern population (Figure 1). A large genetic contribution from Mexican Western populations might be expected for at least 3 reasons. First, the Mexican Western population is likely to be a source population for *I. purpurea* in Mexico, due to its higher diversity and thus larger effective population size (Figure 3). Therefore, all the samples in the Eastern and Southern populations are likely to derive from ancestral populations most similar to the Western population. Second, based on the European migration hypothesis, the early trade routes from Mexico to Spain were from the Valley of Mexico eastwards through Xalapa and on to the port of Veracruz. The relationship between samples in the SE US and the Mexican Eastern population is likely to be more recent, whereas the relationship between samples in the SE US and Mexican Western population is more ancient, so large complete haplotypes and unique SNPs are more likely to be shared between the SE US and Mexican Eastern populations. The shorter haplotype segments used in the genetic assignment analysis are based on partial recombinational independence between locus segments, potentially reflecting more ancient shared identity between the SE US and Mexican Western populations. Third, maize was domesticated ~9000 years ago in the Balsas River Valley (Matsuoka et al. 2002; van Heerwaarden et al. 2011), which is ~300 km southwest of the Western population. Assuming that *I. purpurea* was an early commensal of maize fields, it is likely that the color variants of *I. purpurea* spread throughout Mexico and Central America along with maize culture but may have originated in western Mexico. Presumably there was ample time for gene flow between semidomesticated *I. purpurea* and wild forms as the maize–*Ipomoea* domestication complex diffused outwards.

In addition, substantial trade in the Aztec and earlier eras existed between the Valley of Mexico and regions to the east and south of Mexico, so there was an opportunity for seed transport before the arrival of Europeans. Finally, in modern times horticultural trade may have reintroduced *I. purpurea* to Mexico, with likely escapes from household gardens. So the accessions from around the Valley of Mexico that show high similarity to the SE US population could also have accompanied modern travelers, just as modern cultivars of maize and other improved crops have been reintroduced from the United States to Mexico (Pineyro-Nelson et al., 2009).

Although it is likely that *I. purpurea* first migrated to Europe before being introduced into the SE US, little additional information would be gained from European samples because the time since introduction from Mexico to Europe is known to be very recent. The expected number of new mutation events or of intragenic recombination events is simply too small over a period of 400–500 years to provide any additional resolution. It is the longer time scales in Mexico, together with the population differentiation in Mexico, that provide resolution for inferring origins. There is reason to believe that a more precise geographic origin can be identified because a number of high-frequency haplotypes are found in the SE US that do not occur in the Mexican accessions (e.g., at loci *DFR-B*, *UF3GT*, *FLS*, *F3H*, *IpWRD1*; Figure 5; see Supplementary Figure S4 online). These haplotypes are likely to be present in Mexican source populations. A finer scale geographic sample mesh might identify source populations or regions obscured by the limited resolution of our current sample. In particular, further sampling of the intersection of Eastern and Western populations might provide a more precise geographic definition of the source of the SE US populations.

Supplementary Material

Supplementary material can be found at http://www.jhered.oxfordjournals.org/.

The GenBank numbers for all loci of the SE US accessions are JQ819264–JQ819353; Mexican accessions are JQ819354–JQ819719 and the out-group samples are JQ618023–JQ618032.

Funding

University of Minnesota Grant-in-Aid of Research, Artistry and Scholarship.

Acknowledgments

We thank T. Kono and 2 anonymous reviewers for comments on an earlier version of the manuscript. This work was carried out in part using hardware and software provided by the University of Minnesota Supercomputing Institute.

References

Auld BA, Medd RW. 1987. Weeds: an illustrated botanical guide to the weeds of Australia. Melbourne (Sydney): Inkata Press.

Aylor DL, Price EW, Carbone I. 2006. SNAP: combine and map modules for multilocus population genetic analysis. Bioinformatics. 22:1399–1401.

Baucom RS, Chang SM, Kniskern JM, Rausher MD, Stinchcombe JR. 2011. Morning glory as a powerful model in ecological genomics: tracing adaptation through both natural and artificial selection. Heredity. 107:377–385.

Baucom RS, Mauricio R. 2008. Constraints on the evolution of tolerance to herbicide in the common morning glory: resistance and tolerance are mutually exclusive. Evolution. 62:2842–2854.

Bhangale TR, Stephens M, Nickerson DA. 2006. Automating resequencing-based detection of insertion-deletion polymorphisms. Nat Genet. 38:1457–1462.

Brown BA, Clegg MT. 1984. Influence of flower color polymorphism on genetic transmission in a natural population of the common morning glory, *Ipomoea purpurea*. Evolution. 38:796–803.

Buckler ES, Gaut BS, McMullen MD. 2006. Molecular and functional diversity of maize. Curr Opin Plant Biol. 9:172–176.

Chen H, Morrell PL, Ashworth VE, de la Cruz M, Clegg MT. 2009. Tracing the geographic origins of major avocado cultivars. J Hered. 100:56–65.

Chen H, Morrell PL, Toleno DM, Lundy KE, Clegg MT. 2010. Allele-specific PCR can improve the efficiency of experimental resolution of heterozygotes in resequencing studies. Mol Ecol Resour. 10:647–658.

Clegg MT, Durbin ML. 2000. Flower color variation: a model for the experimental study of evolution. Proc Natl Acad Sci USA. 97:7016–7023.

Clegg MT, Durbin ML. 2003. Tracing floral adaptations from ecology to molecules. Nat Rev Genet. 4:206–215.

Defelice MS. 2001. Tall Morningglory, *Ipomoea purpurea* (L.) Roth - flower or foe? Weed Technol. 15:601–606.

Durbin ML, McCaig B, Clegg MT. 2000. Molecular evolution of the chalcone synthase multigene family in the morning glory genome. Plant Mol Biol. 42:79–92.

Ennos RA. 1981. Quantitative studies of the mating system in two sympatric species of *Ipomoea* (Convolvulaceae). Genetica. 57:93–98.

Ennos RA, Clegg MT. 1983. Flower color variation in the morning glory, *Ipomoea purpurea*. J Hered. 74:247–250.

Epperson BK, Clegg MT. 1986. Spatial-autocorrelation analysis of flower color polymorphisms within substructured populations of morning glory (*Ipomoea purpurea*). Am Nat. 128:840–858.

Epperson BK, Clegg MT. 1987. Frequency-dependent variation for outcrossing rate among flower-color morphs of *Ipomoea purpurea*. Evolution. 41:1302–1311.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 14:2611–2620.

Ewing B, Green P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8:186–194.

Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8:175–185.

Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. Genetics. 155:1405–1413.

Felsenstein J. 2006. Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? Mol Biol Evol. 23:691–700.

Gattepaille LM, Jakobsson M. 2012. Combining markers into haplotypes can improve population structure inference. Genetics. 190:159–174.

Glover D, Durbin ML, Huttley G, Clegg MT. 1996. Genetic diversity in the common morning glory. Plant Species Biol. 11:41–50.

Gonzales AM, Fang Z, Durbin ML, Meyer KKT, Clegg MT, Morrell PL. 2012. Nucleotide sequence diversity of genes involved in floral pigment production in *Ipomoea purpurea*. J Hered. 103:863–872.

Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. Genome Res. 8:195–202.

Halvorson WL, Guertin P. 2003. USGS Weeds in the West project: status of introduced plants in southern Arizona parks. Tucson (AZ): USGS, Sonoran Desert Research Station, University of Arizona.

Hart JP, Brumbach HJ, Lusteck R. 2007. Extending the phytolyth evidence for early maize (*Zea mays ssp mays*) and squash (*Cucurbita sp*) in Central New York. Am Antiq. 72:563–583.

Hill JH. 2001. Proto-Uto-Aztecan: a community of cultivators in Central Mexico? Am Anthropol. 103:913–934.

Hudson RR. 1991. Gene genealogies and the coalescent process. Oxf Surv Evol Biol. 7:1–44.

Hudson RR. 2000. A new statistic for detecting genetic differentiation. Genetics. 155:2011–2014.

Hudson RR, Boos DD, Kaplan NL. 1992. A statistical test for detecting geographic subdivision. Mol Biol Evol. 9:138–151.

Hudson RR, Kaplan NL. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics. 111:147–164.

Huttley GA, Durbin ML, Glover DE, Clegg MT. 1997. Nucleotide polymorphism in the chalcone synthase-A locus and evolution of the chalcone synthase multigene family of common morning glory *Ipomoea purpurea*. Mol Ecol. 6:549–558.

Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 23:1801–1806.

Kimura M, Crow JF. 1964. The number of alleles that can be maintained in a finite population. Genetics. 49:725–738.

Lawson DJ, Hellenthal G, Myers S, Falush D. 2012. Inference of population structure using dense haplotype data. PLoS Genet. 8:e1002453.

Lu Y, Rausher MD. 2003. Evolutionary rate variation in anthocyanin pathway genes. Mol Biol Evol. 20:1844–1853.

Mabberley DJ. 1997. The plant-book: a portable dictionary of the vascular plants. Cambridge (UK): Cambridge University Press.

Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez G J, Buckler E, Doebley J. 2002. A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci USA. 99:6080–6084.

Merrill WL, Hard RJ, Mabry JB, Fritz GJ, Adams KR, Roney JR, MacWilliams AC. 2009. The diffusion of maize to the southwestern United States and its impact. Proc Natl Acad Sci USA. 106:21019–21026.

Miller RE, Rausher MD, Manos PS. 1999. Phylogenetic systematics of Ipomoea (*Convolvulaceae*) based on ITS and waxy sequences. Syst Bot. 24:209–227.

Morrell PL, Clegg MT. 2007. Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. Proc Natl Acad Sci USA. 104:3289–3294.

Morrell PL, Toleno DM, Lundy KE, Clegg MT. 2006. Estimating the contribution of mutation, recombination and gene conversion in the generation of haplotypic diversity. Genetics. 173:1705–1723.

Pineyro-Nelson A, van Heerwaarden J, Perales HR, Serratos-hernandez JA, Rangel A, Hufford MB, Gepts P, Garay-Arroyo A, Rivera-Bustamante R, Álvarez-Buylla ER. 2009. Transgenes in Mexican maize: molecular evidence and methodological considerations for GMO detection in landrace populations. Mol Ecol. 18:750–761.

Piperno DR, Ranere AJ, Holst I, Iriarte J, Dickau R. 2009. Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. Proc Natl Acad Sci USA. 106:5019–5024.

Pluzhnikov A, Donnelly P. 1996. Optimal sequencing strategies for surveying molecular genetic diversity. Genetics. 144:1247–1262.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics. 155:945–959.

Rosenberg NA. 2005. Algorithms for selecting informative marker panels for population assignment. J Comput Biol. 12:1183–1201.

Rosenberg NA, Li LM, Ward R, Pritchard JK. 2003. Informativeness of genetic markers for inference of ancestry. Am J Hum Genet. 73:1402–1422.

Ross-Ibarra J, Tenaillon M, Gaut BS. 2009. Historical divergence and gene flow in the genus Zea. Genetics. 181:1399–1413.

Smith BD. 2006. Eastern North America as an independent center of plant domestication. Proc Natl Acad Sci USA. 103:12223–12228.

Stucky JM. 1985. Pollination systems of sympatric *Ipomoea hederacea* and *I. purpurea* and the significance of interspecific pollen flow. Am J Bot. 72:32–43.

Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics. 105:437–460.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123:585–595.

Thornton K. 2003. Libsequence: a C++ class library for evolutionary genetic analysis. Bioinformatics. 19:2325–2327.

Toleno DM, Durbin ML, Lundy KE, Clegg MT. 2010. Extensive evolutionary rate variation in floral color determining genes in the genus *Ipomoea*. Plant Spec Biol. 25:30–42.

Toleno DM, Morrell PL, Clegg MT. 2007. Error detection in SNP data by considering the likelihood of recombinational history implied by three-site combinations. Bioinformatics. 23:1807–1814.

van Heerwaarden J, Doebley J, Briggs WH, Glaubitz JC, Goodman MM, de Jesus Sanchez Gonzalez J, Ross-Ibarra J. 2011. Genetic signals of origin, spread, and introgression in a large sample of maize landraces. Proc Natl Acad Sci USA. 108:1088–1092.

Villasenor R, Espinosa G. 1998. Catalog of weeds from Mexico. Mexico: UNAM. National Advisory Council on Plant Health. Economic Culture Fund.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. Theor Popul Biol. 7:256–276.

Webster TM, Coble HD. 1997. Changes in the weed species composition of the southern United States: 1974 to 1995. Weed Technol. 11:308–317.

Received January 25, 2013; First decision March 4, 2013; Accepted June 12, 2013

Corresponding Editor: Kenneth Olsen