Journal of Heredity 2012:103(6):863–872
doi:10.1093/jhered/ess059
Advance Access publication October 22, 2012

Nucleotide Sequence Diversity of Floral Pigment Genes in Mexican Populations of *Ipomoea purpurea* (Morning Glory) Accord with a Neutral Model of Evolution

Ana M. Gonzales, Zhou Fang, Mary L. Durbin, Kapua K. T. Meyer, Michael T. Clegg, and Peter L. Morrell

From the Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108 (Gonzales, Fang, and Morrell); and the Department of Ecology and Evolutionary Biology, University of California, Irvine, Irvine, CA 92697 (Durbin, Meyer, and Clegg).

Address correspondence to Peter L. Morrell at the address above, or e-mail: pmorrell@umn.edu.

Abstract

The common morning glory (*Ipomoea purpurea*) is an annual vine native to Central and Southern Mexico. The genetics of flower color polymorphisms and interactions with the biotic environment have been extensively studied in *I. purpurea* and in its sister species *I. nil*. In this study, we examine nucleotide sequence polymorphism in 11 loci, 9 of which are known to participate in a pathway that produces floral pigments. A sample of 30 *I. purpurea* accessions from the native range of Central and Southern Mexico comprise the data, along with one accession from each of the two sister species *I. alba* and *I. nil*. We observe moderate levels of nucleotide sequence polymorphism of ~1%. The ratio of recombination to mutation parameter estimates (ρ/θ) of ~2.5 appears consistent with a mixed-mating system. *Ipomoea* resequencing data from these genic regions are noteworthy in providing a good fit to the standard neutral model of molecular evolution. The derived silent site frequency spectrum is very close to that predicted by coalescent simulations of a drift-mutation process, and Tajima's D values are not significantly different from expectations under neutrality.

Key words: *diversity, mutation, recombination, tropical plants*

It has long been known that biological diversity and species richness increase toward the tropics. Less is known about levels of nucleotide sequence diversity within and among tropical species. A recent resequencing study of the subtropical tree species *Persea americana* (avocado) found that levels of nucleotide sequence diversity were typical of plant species from temperate regions, suggesting that the mutation-drift-selection processes that determine sequence diversity are similar in both temperate and subtropical regions (Chen et al. 2008; Chen et al. 2009). Our goal in this study is to extend resequencing studies to a second subtropical species that differs dramatically in life history, the common morning glory (*Ipomoea purpurea*). We focus on a set of genes that determine floral color.

Ipomoea is a pantropically distributed genus that illustrates the trend of increasing diversity in tropical and subtropical species. *Ipomoea* includes more than 600 species worldwide (Austin and Huaman 1996), with approximately 500 species in tropical and subtropical regions of the Americas. The common morning glory, *I. purpurea*, is an annual self-compatible vine native to Central and Southern Mexico characterized by large showy flowers. Although not native to the region, the common morning glory is also found in the Southeastern United States (Halvorson and Guertin 2003). The timing and routes of introduction into the Southeastern United States are uncertain (Glover et al. 1996; Clegg and Durbin 2000), but it is possible that European colonists introduced the species into the Southeast. In the United States, *I. purpurea* is classified among the 11 most "troublesome" weeds (Webster and Coble 1997). In central Mexico, *I. purpurea* populations tend to occur from ~800 m to ~1800 m elevation in contrast to the sister species, *I. nil*, that occurs from sea level to

~1000 m (unpublished data). As suggested by the common name, morning glory flowers open early in the day and are available for pollination for a few hours before closing and abscising from the plant. The common morning glory is likely to have been partially domesticated by pre-Hispanic civilizations, perhaps in association with maize culture (Clegg and Durbin 2003). Presumably, Neolithic plant domesticators prized the diverse flower color mutations that appeared occasionally, and these were selected and propagated. In cultivation, common morning glory flowers are found in diverse colors including white, pink, blue, and dark blue (purple). It is now known that many of the underlying mutations in *I. purpurea* are induced by transposable elements and may occur at high frequencies (Epperson and Clegg 1992; Habu et al. 1998; Clegg and Durbin 2003).

Ipomoea has been extensively studied to understand the genetic bases of flower color differences (Barker 1917; Ennos and Clegg 1983). Most flower colors are the result of the presence of anthocyanin pigments or other flavonoid compounds in floral tissue. Anthocyanins are the final product of the flavonoid biosynthesis pathway. This pathway has provided a model system for analyzing the processes of selection at the molecular, biochemical, and phenotypic levels (Schoen and Clegg 1985; Epperson and Clegg 1987; Glover et al. 1996; Rausher et al. 1999; Clegg and Durbin 2003). *Ipomoea purpurea* and *I. nil* are the two most studied *Ipomoea* species for flower color polymorphism. Studies of *I. purpurea* flower color variation suggest that color can influence pollinator behavior, preferences, or even the attraction of new pollinators; therefore, color changes can affect reproductive success and fitness (Barker 1917; Brown and Clegg 1984; Schoen and Clegg 1985; Epperson and Clegg 1987; Rausher et al. 1993).

Previous studies of genes involved in flower color polymorphism have examined relative rates of evolution at a phylogenetic scale. For example, a study in *I. purpurea*, *Antirrhinum majus*, and *Zea mays* compared nonsynonymous substitution rates in 6 genes from the anthocyanin pathway. This study determined that upstream genes evolve more slowly than downstream genes, leading to speculation that the pattern of constraints on a gene depends on participation in other pathways (Rausher et al. 1999). Another sequencing study comparing 4 genes in the anthocyanin biosynthesis pathway in 9 species across the genus *Ipomoea* found that the majority of the phenotypic differences between species were due to changes in regulation of gene expression (Durbin et al. 2003). A recent study by Toleno et al. (2010) included the analysis of rates of evolution in 6 genes from the anthocyanin pathway across 19 species of *Ipomoea*. This study reported only a weak correlation between pathway position and evolutionary constraint, and Toleno et al. (2010) concluded that there is little evidence for positive selection on structural genes in the anthocyanin biosynthesis pathway in a broad sample of *Ipomoea* species.

The majority of studies of *I. purpurea* have focused on weedy, naturalized populations in the Southeastern United States rather than the native range of *I. purpurea* in Mexico, which limits insight into the ecological circumstances and genetic history that have driven the evolution of the species (Clegg and Durbin 2000). To address this omission, the present study analyzes data from the resequencing of 11 genes from 30 accessions of *I. purpurea* sampled from the species' native range in Central and Southern Mexico. One Mexican accession of *I. nil* and an accession of *I. alba* are included as out group samples. The study of nucleotide polymorphism in *Ipomoea* populations from the natural subtropical range allows us to compare the roles of selection and drift in determining sequence diversity; the importance of recombination relative to mutation in generating haplotype diversity; and the diversity in subtropical and temperate species.

Materials and Methods

Plant Materials

Our sample includes 30 accessions of *I. purpurea* from across the native range of the species in Mexico (see supplementary Table S1 online). Single accessions are used to represent wild populations in Central and Southern Mexico including populations from Chiapas, Estado de Mexico, Michoacán, Morelos, Oaxaca, and Veracruz. Seeds from the wild isolates were collected in Mexico and germinated in a greenhouse in California. All the individuals in this study have purple flowers (wild type) with minimal variation in color intensity. We also include accessions of *Ipomoea alba* and *I. nil* to permit inference of the ancestral state of mutations at each locus.

Loci Resequenced

Genes known to be involved in the production of flower color in *I. purpurea* were resequenced. This includes 6 structural genes in the anthocyanin biosynthesis pathway that contribute to the production of anthocyanins or flavonols. Listed earlier to later in the anthocyanin biosynthesis pathway, the loci resequenced here include *CHS-D, CHS-E, F3H, FLS, DFR-B*, and *UF3GT* (see supplementary Table S2 online), and 4 transcription factors, 3 of which have been reported to regulate levels of expression of genes in the anthocyanin biosynthesis pathway: *IpMyb1, IpbHLH1*, and *IpWDR1*. *IpMyb4* is a member of the R2R3 MYB family of regulatory genes, but its function is currently undetermined although its expression is limited to floral tissue (unpublished data). We also included *ALS*, a gene from a primary pathway for leucine biosynthesis, and used for relative rates comparisons by Toleno et al. (2010).

DNA Sequencing and Assembly

Amplicons for each locus and accession were used for direct Sanger sequencing with Big Dye version 3.1 sequencing chemistry. Amplicons were sequenced with both the initial amplification primers and internal primers to obtain sequence quality of phred ≥ 20 for both the forward and reverse strands. Sequence reads were assembled using phred for base-calling and phrap for assembly (Ewing and Green 1998). Reads were visualized with consed (Gordon et al.

1998). PolyPhred (Nickerson et al. 1997) was used to detect single nucleotide polymorphisms (SNPs) and insertion–deletion (indel) mutations. When a sample was determined to be heterozygous for a particular amplicon, the sample was cloned using Qiagen pDrive Cloning Vector. At least 3 clones per haplotype were used for experimental phasing. Thus all data reported are based on fully experimentally resolved haplotypes. We tested the accuracy of inferred haplotypes using the program Error Detection Using Triplets (Toleno et al. 2007).

Sequence Analysis

Sequence Diversity Estimation, Tests of Neutrality

Descriptive statistics and nucleotide polymorphism estimates were calculated using tools from the libsequence C++ library (Thornton 2003). Reported statistics include haplotype number (*h*), the number of segregating sites (*S*), and 2 estimates of the mutation parameter, θ_W and θ_π. Both θ_W and θ_π estimate $4N_e\mu$, where N_e is the effective population size and μ is the mutation rate per base pair per generation, but θ_W (Watterson 1975) is based on observed segregating sites, and θ_π (Tajima 1983) is based on average pairwise differences between samples. The normalized difference between these 2 estimators, Tajima's D, is used as a test of deviation from neutrality under a standard coalescent model (Tajima 1989). We also report Fay and Wu's H (Fay and Wu 2000) a test of neutrality that incorporates information on the derived versus ancestral state of mutations. Relative levels of diversity in various partitions of the data, for example, introns versus exons and silent versus replacement sites at each locus were estimated using the program polydNdS from libsequence (Thornton 2003).

We also assessed departures from neutrality attributable to selection by performing a Hudson-Kreitman-Aguade (HKA) test (Hudson et al. 1987), using a maximum likelihood implementation in the program MLHKA (Wright and Charlesworth 2004). The HKA test evaluates the ratio of polymorphism among individuals within a species relative to divergence from an out group and tests for differentiation from this ratio at putatively neutrally evolving loci (Hudson et al. 1987). MLHKA permits the testing of individual loci in a multilocus framework. The divergence time parameter between *I. purpurea* and the out group (*I. alba* or *I.nil*) used here was 3 million years (Clegg and Durbin 2000) scaled by $2N_e$ generations. We estimated N_e (effective population size) from $\theta_W = 4N_e\mu$ using a neutral mutation rate of $\mu = 6.5 \times 10^{-9}$ (Gaut et al. 1996) and the mean value of θ_W from the empirical data set. The program was run with 3 different models, 1) assuming that all loci are evolving under neutrality, 2) assuming that all loci are subject to selection, and 3) assuming each locus individually could be subject to selection. The significance of the test ($P < 0.05$) was assessed using a likelihood ratio test, where twice the difference of the log likelihood between the 2 models is approximately χ^2 distributed. Each model was run for 100 000 chains, using a random seed number. Two runs per model were performed to estimate maximum likelihood values.

We report 2 estimates of the recombination rate parameter $\rho = 4N_er$, where *r* is the recombination rate per base pair per generation. The first estimation is based on a maximum composite likelihood approach implemented in maxhap (referred to as ρH; Hudson 2001). This approach was also used to test for the presence of gene conversion at each locus, by estimating the relative contribution of gene conversion to crossing over (*f*). The second approach uses Approximate Bayesian Computation (ABC; Beaumont et al. 2002) based on linear regression implemented in the ABCreg software (Thornton 2009). We jointly estimated ρ_{ABC} and the population mutation rate per bp (θ_{ABC}). We used 5 summary statistics for each locus: number of segregating sites (*S*); pairwise differences per locus (θ_π); number of haplotypes (*h*); a minimum bound on the number of recombination events based on the 4-gamete test (R_m; Hudson and Kaplan 1985); and an estimate of the number of recombination events based on the difference between *h* and *S* ($h - S - 1 = R_h$; Myers and Griffiths 2003). Coalescent simulations were generated based on uniform distributions of ρ and θ, with the descriptive statistics detailed above compared with empirical values for each locus. Priors were chosen to bracket point estimates from ρ_W and θ_H. Priors were further adjusted if posterior distributions were constrained by the priors. The posterior distribution of the per-site population mutation rate, the recombination rate, and the ratio of recombination to mutation for each locus were determined based on linear regression with a tangent transformation.

As an additional means of assaying for the presence of homologous gene conversion at each locus, we used a pattern matching approach (Padhukasahasram et al. 2004) to calculate the proportion of "pattern *a*" and "pattern *d*" (Padhukasahasram et al. 2004; Morrell et al. 2006) present at each locus. These 2 patterns are defined by SNP configurations; pattern *a* includes all 4-gamete combinations between external sites (defined as sites A and C from a set of sites in linear order defined as ABC) but does not include all 4 gametic combinations for sites A with B or B with C. Pattern *d* is defined as including external sites A and D, from an ABCD configuration, which contain all 4-gametic combinations along with internal sites B and C that do not. Under an infinite site assumption, the patterns cannot be explained by a single crossover but could result from a single gene conversion event, with the additional stipulation that pattern *d* involves 2 SNPs within the pattern. This has been identified as a clear conversion (Plagnol et al. 2006; Song et al. 2006) because the pattern cannot arise from a back mutation at a single nucleotide site.

Derived Site Frequency Spectrum

The derived (or unfolded) nucleotide site frequency spectrum (SFS) was calculated using a Perl version of the software SoFoS (available at www.rilab.org, version of January 2012) to estimate the number of observed derived alleles at each site. We used hypergeometric scaling (Nielsen et al. 2004) to rescale the sample size to the smallest chromosome number among all loci. The ancestral state for each SNP was inferred relative to accessions of *I. alba* or *I.nil* using a simple parsimony criterion. The SFS was calculated separately for silent and replacement polymorphisms.

Coalescent simulations were generated using ms (Hudson 2002) to compare the empirical derived SFS to that expected under a neutral model. We used 10 000 simulations based on the average sample size, and median values of θ_W and ρ_H across loci. A Pearson Goodness-of-fit test was used to test the fit of the empirical SFS to simulations.

Results

Nucleotide Sequence Polymorphism

Resequencing of 30 accessions of *I. purpurea* resulted in an average aligned length per locus of 1001 bp from 11 loci. This includes an average of 745 bp of exonic and 134 bp of intronic sequence per locus. For *CHS-D*, *CHS-E*, and *IpMyb4*, all sequence is from a single exon. Observed heterozygosity averaged over loci was 10%; 16 accessions were heterozygous for at least 1 amplicon, with an average of 3 detectable heterozygotes per locus. To account for the mixed-mating system, we included both haplotypes from heterozygous individuals in the sample. Thus, the sample size for each locus ranged from 31 to 36 chromosomes depending on the number of heterozygous individuals.

The average number of segregating sites (*S*) per locus was 32. The aligned sequence includes indel polymorphisms, including 2 indels that occurred within exons while maintaining the reading frame at the locus. Seven sites had 2 or more mutations; these sites and indel polymorphisms were maintained in the alignment but excluded from the diversity calculations.

Nucleotide diversity at each locus is shown in Table 1. The most diverse locus in the sample is *DFR-B* with $\theta_\pi = 22.4 \times 10^{-3}$, and the least diverse is *IpbHLH1* with $\theta_\pi = 3.0 \times 10^{-3}$. Mean and median θ_W across loci were 11.7 and 8.4×10^{-3} respectively, with mean and median θ_π of 10.5 and 7.9×10^{-3}. All loci except *F3H* have lower diversity at nonsynonymous than synonymous sites (θ_π nonsyn/syn), consistent with strong purifying selection.

Tajima's D estimates at each locus do not differ significantly from zero, consistent with expectations under neutrality (Table 1). The mean Tajima's D is −0.45, with the most positive value of 0.15 at *F3H,* and *UF3GT* the most negative at −1.18 (see supplementary Figure S1A online). When the data were partitioned into exons and introns, we observed that *ALS* and *IpMYB1* have negative values for exons, but positive values for introns (see supplementary Figure S1B and S1C online). The excess of negative Tajima's D in exons is consistent with an excess of rare nonsynonymous changes in exons observed at these loci, as might be expected under selection–mutation balance. This effect is also observed in the unfolded SFS (Figure 1). The Fay and Wu's H statistic (Table 1), which is less sensitive to demographic effects and more sensitive to positive selection (Fay and Wu 2000), has a mean value close to zero (-0.8×10^{-3}, maximum = 9.11×10^{-3}, and minimum = -16.66×10^{-3}) also consistent with no direct evidence of positive selection based on polymorphism data.

MLHKA tests were significant at $P < 0.05$ level for *ALS, FLS*, and *IpbHLH1* (Table 2). The degree to which diversity differed from neutral expectations is measured with the parameter *k*, where values >1 indicate an elevation of diversity relative to divergence. *ALS* had the most elevated value with $k = 4.8$. When we applied a Bonferroni correction for multiple testing ($\alpha = 0.004$), the analysis of departures from neutrality due to selection remain significant, but the deviation observed in *ALS* is not significant ($P < 0.004$).

Recombinational Diversity

Recombination events are evident at all loci with average $R_m = 6.7$ (Table 3). Three loci have very large R_m values, with $R_m \geq 11$ and very large values of R_h. Parametric estimates of recombination rates using a coalescent-based estimator reveals per base pair ranges of ρ_H from 5.9×10^{-3} to 123.4×10^{-3} with a median of $\rho_H = 31.2 \times 10^{-3}$.

Table 1 Descriptive statistics and nucleotide diversity for 11 loci in *I. purpurea*

Gene	#Sites (bp)	Exon (bp)	H_o	*S*	Singletons	*h*	Haplotype diversity	$\Theta_W \times 10^{-3}$	$\Theta_\Pi \times 10^{-3}$	Tajima's *D*	$\Theta_{\Pi syn} \times 10^{-3}$	Θ_Π non-syn/syn	Fay and Wu's $H \times 10^{-3}$
CHS-D	903	903	0.12	31	14	21	0.96	8.40	6.15	−0.95	25.93	0.01	3.79
CHS-E	894	894	0.17	39	14	19	0.92	10.79	10.45	−0.11	46.16	0.00	−8.61
DFR-B	1247	408	0.06	77	16	18	0.95	24.98	22.42	−0.39	16.70	0.15	6.32
UF3GT	1088	1008	0.14	67	28	23	0.95	15.99	10.89	−1.18	23.31	0.19	2.54
F3H	961	776	0.06	29	8	19	0.92	7.64	7.95	0.15	5.65	1.73	−14.75
FLS	1113	779	0.03	32	12	17	0.87	7.64	6.63	−0.48	11.06	0.22	−16.66
ALS	808	732	0.14	22	9	19	0.94	6.94	5.44	−0.74	14.33	0.07	0.64
IpHLH1	1230	654	0.06	18	6	11	0.81	3.64	3.00	−0.60	4.99	0.07	0.73
IpWDR1	1145	1011	0.17	32	9	24	0.97	6.75	4.89	−0.97	14.68	0.03	1.07
IpMYB4	489	489	0.06	27	9	12	0.86	16.23	17.85	0.36	15.83	0.85	5.98
IpMYB1	1141	540	0.03	54	19	21	0.96	19.96	19.56	−0.07	14.54	0.40	9.11
Average	1001.7	744.9	0.10	38.9	13	18.5	0.92	11.72	10.48	−0.45	17.56	0.34	−0.89

Symbols as defined in the text. Abbreviations: H_O, percentage of observed heterozygosity; S, segregating sites; h, haplotype number.

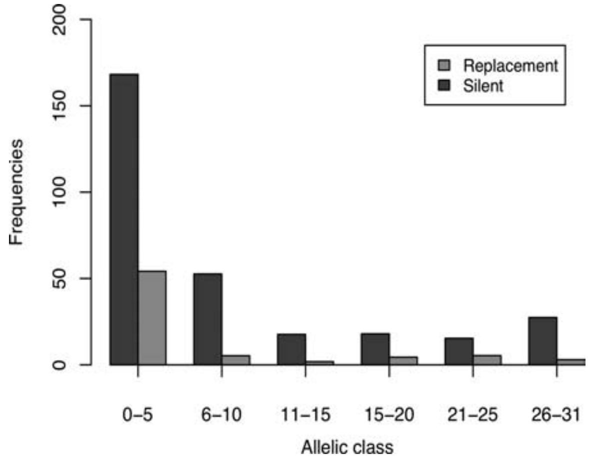

Figure 1. The derived site frequency spectrum for 11 loci in *I. purpurea* at silent and replacement sites.

Table 2 Maximum likelihood HKA analysis of selection at all loci

	ln *L*	Likelihood ratio statistic (df)	*k*	*P* value (<0.05)
Neutral	−89.25		—	
All loci	−75.96	26.58 (10)	—	0.003
CHS-D	−88.75	0.99	0.53	0.319
CHS-E	−88.52	1.45	0.55	0.229
DFR-B	−88.77	0.96	1.55	0.328
UF3GT	−89.25	0.00	0.93	0.970
F3H	−88.41	1.67	0.54	0.196
FLS	−80.65	17.19	2.61	<0.0001
ALS	−86.04	6.42	4.80	0.011
IpHLH1	−85.12	8.25	0.20	0.004
IpWDR1	−89.27	0.05	1.11	0.821
IpMYB4	−89.22	0.05	1.20	0.816
IpMYB1	−89.14	0.22	0.75	0.643

Abbreviations: ln *L*, log-likelihood value for each model; k, the degree to which diversity is increased or decreased relative to neutral expectations.

The ratio of ρ_H to θ_w is 4.4. When the recombination rate was estimated using ABC regression (Figure 2, Table 3), the mean ρ_{ABC} was 17.5 × 10^{-3} and the mean θ_{ABC} was 9.53 × 10^{-3}. Point estimates of ρ_{ABC}/θ_{ABC} average 2.52 over all 11 loci, with mean lower and upper 95% confidence intervals of 0.88 and 5.33, respectively. Both estimators suggest a major contribution of recombination to haplotype diversity at these loci.

Estimating Gene Conversion Rates

Using Hudson's estimator for gene conversion (*f*), the majority of loci demonstrate evidence of gene conversion with $f > 0$ (Table 4). All but one locus have pattern *a* combinations present, consistent with homologous gene conversion or double crossover. In addition, 10 loci demonstrate the presence of pattern *d* within the locus, with 2 or more SNPs within a tract, suggesting a gene conversion event, a pattern known as a clear conversion.

Diversity at Regulatory versus Structural Genes

We divided the data into regulatory and structural genes to ask how these 2 categories affect diversity. These data do not reveal any notable differences in diversity between regulatory and structural genes (mean θ_π at all sites is 0.01 in both cases). However, when the data were partitioned into silent and replacement sites (see supplementary Figure S2 online), regulatory genes have slightly more replacement polymorphisms than structural genes.

Discussion

Three key results of this study are the following: (1) The data provide little evidence of positive selection acting on flower color–determining genes, despite the role of flower color in reproductive success. A neutral model with purifying selection best accounts for the patterns of polymorphism in native populations of *I. purpurea* in Mexico. (2) Recombination and

Table 3 Descriptive statistics for recombinational diversity

Gene	Wall's *B*	R_h	R_m	$\rho_H \times 10^{-3}$	ρ_H/θ_W	$\rho_{ABC} \times 10^{-3}$	ρ_{ABC}/θ_{ABC}
CHS-D	0.03	11	2	14.80	1.76	10.53	1.11
CHS-E	0.08	21	11	98.13	9.09	11.69	3.44
DFR-B	0.04	60	12	15.03	0.6	14.63	0.83
UF3GT	0.13	45	7	31.23	1.95	15.65	1.02
F3H	0.00	11	8	32.07	4.2	38.82	5.30
FLS	0.10	16	8	37.55	4.92	29.54	4.12
ALS	0.10	4	2	31.96	4.61	22.79	3.18
IpHLH1	0.06	8	2	8.91	2.45	6.27	2.04
IpWDR1	0.03	9	6	123.43	18.3	14.99	2.29
IpMYB4	0.22	16	3	5.95	0.37	20.27	1.50
IpMYB1	0.06	34	13	11.62	0.58	31.24	2.82
Average	0.08	21.4	6.7	37.33	4.44	19.69	2.52

Recombination and mutation rate per base pair. Recombination is estimated using a maximum likelihood approach ρ_H and approximate Bayesian computation approach (ρ_{ABC}). R_m and R_h are estimates of the number of observed recombination events.

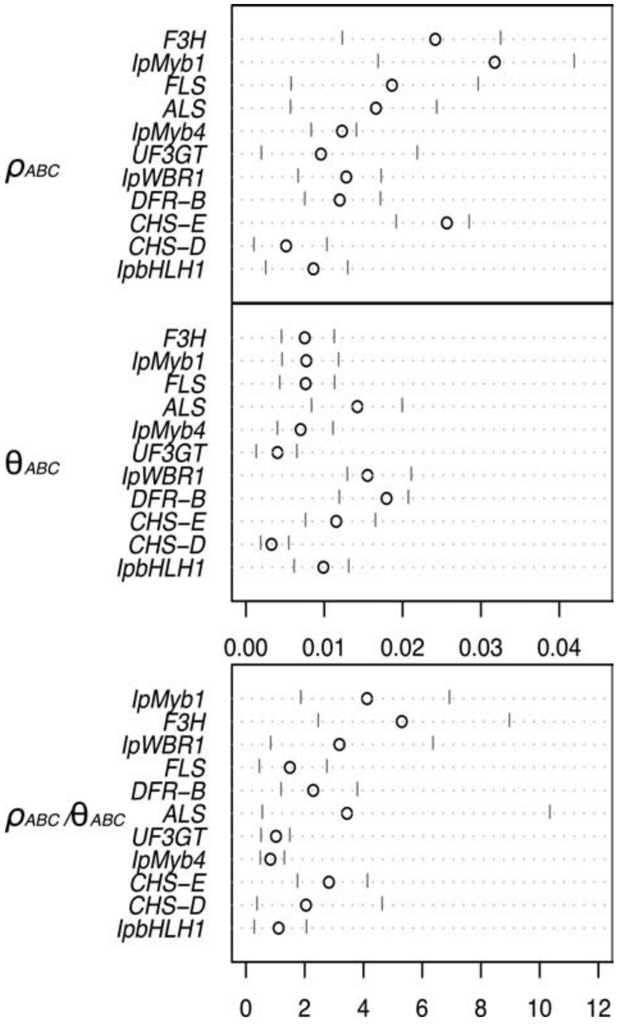

Figure 2. Joint estimates of population recombination and mutation rate based on coalescent simulation and approximate Bayesian computation. Estimates shown are ρ_{ABC} and θ_{ABC} and ρ_{ABC}/θ_{ABC} for 11 loci in *I. purpurea*. Point estimates are displayed as circles, with the 95% confidence interval depicted as vertical bars.

Table 4 The proportion of sites in patterns a and d for 11 loci

Gene	Pattern *a*	Pattern *d*
CHS-D	13.61	2.54
CHS-E	9.30	1.30
DFR-B	11.60	1.60
UF3GT	14.10	2.30
F3H	21.50	4.60
FLS	33.30	5.40
ALS	6.15	0.49
IpHLH1	0.00	0.00
IpWDR1	19.20	3.10
IpMYB4	3.90	0.40
IpMYB1	11.60	2.90

Values shown are multiplied by 10^{-3}.

gene conversion play a larger role than mutation in generating haplotypic diversity. And (3) statistics of population diversity reveal no substantial differences between tropical and temperate species, suggesting comparable effective population size, mutation rate, and recombination parameters. However, existing comparative data, particularly for tropical and subtropical species, are still quite limited.

It is useful to begin with a brief consideration of the haplotype data that provide the basis for this and comparable studies. Inferring haplotypes from sequence data is complicated by heterozygosity within individuals and the diploid (and in some cases polyploid) nature of most eukaryotic species. Simply identifying SNPs does not reveal the arrangement of SNPs on a chromosome. Haplotypes can be resolved either by computational (Clark 1990; Stephens et al. 2001) or by experimental approaches, including cloning or allele-specific polymerase chain reaction (cf. Chen et al. 2010). The experimental approach applied here involves cloning combined with an LD-based error detection approach (Toleno et al. 2007) and iterative steps to improve haplotype accuracy (Chen et al. 2010). The careful, but often painstaking, determination of haplotypes provides an extra dimension of information, especially for estimations of local patterns of linkage disequilibrium and the relative contribution of recombination and mutation to haplotypic diversity. It also ensures that the resulting analyses are as accurate as possible within the limits of sampling error. As a consequence, the present data permit some analyses that are not possible from SNP genotyping data alone and provide a degree of precision that may not always be attained even in much larger SNP studies.

The diversity analyses for both synonymous and nonsynonymous changes at the level of the species, *I. purpurea*, are congruent with the rate and gene rank order of those observed at the level of the *Ipomoea* genus by Toleno et al. (2010) for the 4 structural genes that are common to both studies (*ALS, CHS-D, DFR-B,* and *UF3GT*). From this comparison, we conclude that the evolution of these 4 genes has been largely governed by a neutral process with purifying selection over roughly 40 million years. This consistency is noteworthy in light of the involvement of *DFR-B* and *UF3GT* genes in a biosynthetic pathway that determines the reproductively important floral pigmentation trait, and the frequent shifts in flower color among species in the genus. In this context, it is important to note that Durbin et al. (2003) found that regulatory rather than structural gene change was implicated in most shifts in flower color expression at the genus level in *Ipomoea*.

The current finding of similar rates of diversity in structural and regulatory genes contrasts with that observed by Rausher et al. (1999), who concluded that regulatory genes are less constrained than structural genes and thus may play a more important role in mediating adaptive evolution. These conflicting results may be explained by the differences of the time scale spanned in the 2 studies. The study of Rausher et al. (1999) sampled deeply divergent taxa, resulting in comparisons across much of the history of flowering plants, whereas this study measures diversity within a species, thus

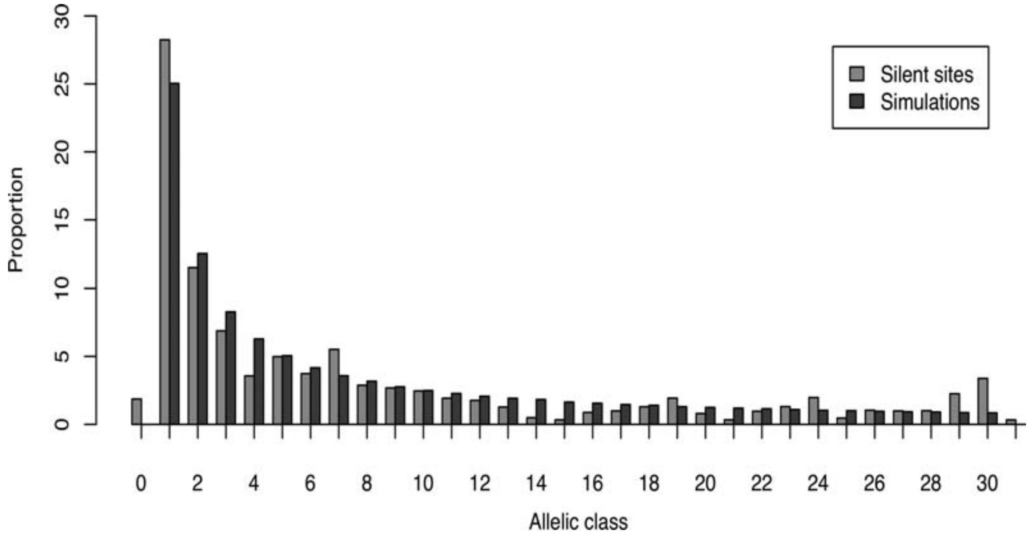

Figure 3. The derived site frequency spectrum at silent sites from 11 loci (light gray) relative to coalescent simulations (dark gray).

reducing the time scale and therefore the potential to observe differences among classes of genes.

Tests for Deviations from Neutrality

Tajima's D for all loci is between −1.5 and +1 (see supplementary Figure S1 online), consistent with expectations under neutrality. When the data are partitioned into exons and introns we observe more negative values of Tajima's D for exonic regions, consistent with the observation of low-frequency nonsynonymous changes. An excess of low-frequency nonsynonymous changes is frequently observed in plant resequencing studies, and has been attributed to the reduced efficacy of purifying selection in small local populations (Cummings and Clegg 1998). As already noted, *I. purpurea* often occupies recently disturbed areas, with populations in Mexico frequently consisting of a relatively small number of twining or sprawling plants in a spatially limited site.

The observed SFS (Figure 1) is also consistent with a drift-mutation process. Comparisons to a simulated data set (Figure 3) based on a coalescent model revealed that the genes involved in anthocyanin biosynthesis are evolving in a manner that is quite similar to expectations under a standard neutral model, with a slight excess of rare silent changes.

The MLHKA test indicates that 3 loci depart from neutral expectations. For these 3 loci, however, polymorphism-based tests do not detect significant departures from neutrality. The *ALS* and *FLS* loci show the largest deviation from neutral expectations, but no reduction in diversity or skew in the SFS, suggesting that selective effects on these loci were not sufficiently strong or recent to be detectable based on standing variation.

Overall, our results indicate that *I. purpurea* provides a relatively good fit to the standard neutral model of evolution with purifying selection. It is not necessary to invoke recent adaptive selection or recent demographic changes within the species history (e.g., population expansion) to explain observed patterns of polymorphism.

Mixed-Mating System and Haplotype Diversity

At the majority of loci, recombination (crossover and some contribution of double crossover or gene conversion) contributes more to total diversity than mutation, with average ρ_H/θ_W of 4.4 and the jointly estimated ratio of $\rho_{ABC}/\theta_{ABC} = 2.5$. From this, we conclude that the role of recombination in the formation of new haplotypes by recombining existing mutations is more important in generating haplotype diversity than mutation. The ratio of $\rho_H/\theta_W = 4.4$ contrasts with recombination rates previously estimated in other subtropical and temperate species, which revealed that recombination plays a more limited role in overall diversity. This is the case for wild avocado (*P. americana*) with $\rho/\theta = 0.8$ (Chen et al. 2008), wild barley (*Hordeum vulgare* ssp. *spontaneum*) with $\rho/\theta = 1.5$ (Morrell et al. 2006), and *A. thaliana* with $\rho/\theta = 0.05$ (Nordborg et al. 2005). Both wild barley and *A. thaliana* are predominantly self-fertilizing species. Although avocado is thought to be an outcrossing species, it is capable of high rates of self-fertilization (Kobayashi et al. 2000).

In contrast, outcrossing rates for *I. purpurea* estimated using allozyme markers range from 0.48 for white flowers to 0.73 for blue and pink flowers (Brown and Clegg 1984). There are no estimates of outcrossing rates from native Mexican populations, but Mexican populations tend to be widely scattered with relatively few individuals per population (MT Clegg, personal observation), suggesting the potential for higher levels of inbreeding in Mexico. Despite this, when taken as a whole, these results appear to be consistent with a relatively high rate of recombination in *I. purpurea*. However, depending on the estimator used, the relative contribution of recombination to diversity in *I. purpurea* appears

Table 5 Diversity of temperate and subtropical species with different mating systems

	Species	$\theta_\pi \times 10^{-3}$	References
Mixed mating	*Eichhornia paniculata*ST	5.2	Ness et al. (2010)
Outcrossing	*Capsella grandiflora*^{T}	7.0	Onge et al. (2011)
	Helianthus annuus^{T}	12.8	Liu and Burke (2006)
	Zea mays ssp. parviglumisST	16.9	Ross-Ibarra et al. (2009)
	Solanum peruvianumST	14.6	Arunyawat et al. (2007)
	Persea americana^{ST,P}	6.6	Chen et al. (2008)
	Cryptomeria japonica^{T,P}	2.5	Kado et al. (2003)
	Populus tremula L., Salicaceae^{T,P}	11.1	Ingvarsson (2005)
	Pseudotsuga menziesii^{T,P}	6.5	Krutovsky and Neale (2005)
Self-fertilizing	*Hordeum vulgare ssp. spontaneum*^{T}	7.4	Morrell et al. (2005)
	Arabidopsis thaliana^{T}	5.0	Aguade (2001)
	Capsella rubella^{T}	4.0	Onge et al. (2011)
	Carthamus palaestinus^{T}	5.1	Chapman and Burke (2007)
	Capsicum annuumST	2.8	Aguilar-Melendez et al. (2009)

Abbreviations: ST, subtropical species; T, tropical species; P, perennial.

slightly lower than that estimated for the outcrossing species teosinte (*Z. mays* ssp. *parviglumis*, the maize progenitor) $\rho/\theta = 4.5$ (Wright et al. 2005).

Nucleotide Sequence Diversity Comparison: Temperate versus Subtropical Plant Species

It has long been known that species diversity increases from temperate to tropical regions of the globe. The genus *Ipomoea* provides a good illustration of this trend, with more than 500 species in Mexico alone. Studies of ribosomal DNA restriction fragment variation and gene sequence data from the chalcone synthase (*CHS*) family (Glover et al. 1996; Huttley et al. 1997) reveal a reduction in levels of variation in *I. purpurea* from the Southeastern United States where the species was introduced, relative to the subtropical populations in Mexico. A review of allozyme polymorphism (Hamrick and Godt 1990) suggested that mating system and geographic range (endemic, regional, or widespread distribution) are the most important correlates with differences in genetic diversity within and between species, where predominantly outcrossing species have more diversity than species with either selfing or mixed-mating systems. However, it is important to confirm these allozyme polymorphism–based trends with resequencing studies that can partition diversity into coding and noncoding portions of the genome and where more accurate estimates of population processes can be obtained.

Ipomoea purpurea has a mixed-mating system, but is characterized by moderate levels of nucleotide sequence diversity ($\theta_\pi = 10.5 \times 10^{-3}$). We compare these results with diversity estimates of annual and perennial species from subtropical and temperate regions, with various mating systems presented in Table 5. The comparisons show that *I. purpurea* has similar rates of diversity to tropical and temperate outcrossing species, but it has higher diversity than the majority of tropical and temperate selfing species. Moreover, these comparisons also show that tropical species have similar levels of genetic variation relative to temperate species, as was suggested by Chen et al. (2008). We observed various levels of variation between species with different mating systems and population census sizes. Selfing species are generally less diverse than outcrossing species. Lower diversity in the tree species can be explained by a reduced population size owing to small, widely scattered, and fragmented populations. The same explanation may hold for the moderate and low levels of diversity observed in *I. purpurea* and *Eichhornia paniculata,* both mixed-mating species, where the species effective population size is modest (estimated here to be 450 346 and 200 154, respectively, based on $\theta_w = 4N_e\mu$, as explained in the Materials and Methods section).

At present, the data are extremely limited and do not permit broad generalizations, but it is noteworthy that both *Persea americana* and *I. purpurea*, two subtropical species with quite different life histories, do not display levels of genetic diversity that are markedly different from temperate plant species. Consequently, estimates of effective population size, mutation, and recombination do not indicate any clear distinction between subtropical and temperate plant species, with respect to fundamental evolutionary processes.

Supplementary Material

Supplementary material can be found at http://www.jhered.oxfordjournals.org/. The sequences of the 11 loci from *Ipomoea purpurea* were submitted to NCBI and include the following accession: JQ819354–JQ819719. The outgroup accession numbers are JQ618023–JQ618032.

Funding

University of Minnesota Grant-in-Aid of Research, Artistry & Scholarship; University of Minnesota Supercomputing Institute.

Acknowledgments

The authors thank Beau Miller for technical assistance, and Thomas Kono, Tanja Pyhäjärvi, and Michael Kantar for comments on an earlier version of the manuscript.

References

Aguadé M. 2001. Nucleotide sequence variation at two genes of the phenylpropanoid pathway, the FAH1 and F3H genes, in *Arabidopsis thaliana*. Mol Biol Evol. 18:1–9.

Aguilar-Meléndez A, Morrell PL, Roose ML, Kim SC. 2009. Genetic diversity and structure in semiwild and domesticated chiles (*Capsicum annuum*; Solanaceae) from Mexico. Am J Bot. 96:1190–1202.

Arunyawat U, Stephan W, Städler T. 2007. Using multilocus sequence data to assess population structure, natural selection, and linkage disequilibrium in wild tomatoes. Mol Biol Evol. 24:2310–2322.

Austin DF, Huaman Z. 1996. A synopsis of *Ipomoea* (Convolvulaceae) in the Americas. Taxon. 45:3–38.

Barker EE. 1917. Heredity studies in the morning-glories (*Ipomoea purpurea (L.) Roth*). editor. Bull. Agric. Exp. Sta. Cornell Univ. 392:3–38.

Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian computation in population genetics. Genetics. 162:2025–2035.

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

Chapman MA, Burke JM. 2007. DNA sequence diversity and the origin of cultivated safflower (*Carthamus tinctorius L.; Asteraceae*). BMC Plant Biol. 7:60.

Chen H, Morrell PL, de la Cruz M, Clegg MT. 2008. Nucleotide diversity and linkage disequilibrium in wild avocado (*Persea americana* Mill.). J Hered. 99:382–389.

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.

Clark AG. 1990. Inference of haplotypes from PCR-amplified samples of diploid populations. Mol Biol Evol. 7:111–122.

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.

Cummings MP, Clegg MT. 1998. Nucleotide sequence diversity at the alcohol dehydrogenase 1 locus in wild barley (*Hordeum vulgare* ssp. *spontaneum*): an evaluation of the background selection hypothesis. Proc Natl Acad Sci USA. 95:5637–5642.

Durbin ML, Lundy KE, Morrell PL, Torres-Martinez CL, Clegg MT. 2003. Genes that determine flower color: the role of regulatory changes in the evolution of phenotypic adaptations. Mol Phylogenet Evol. 29:507–518.

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

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

Epperson BK, Clegg MT. 1992. Unstable white flower color genes and their derivatives in the morning glory. J Hered. 83:405–409.

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

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

Gaut BS, Morton BR, McCaig BC, Clegg MT. 1996. Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. Proc Natl Acad Sci USA. 93:10274–10279.

Glover D, Durbin ML, Huttley G, Clegg MT. 1996. Genetic diversity in the common morning glory. Plant Spec Biol. 11:41–450.

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

Habu Y, Hisatomi Y, Iida S. 1998. Molecular characterization of the mutable flaked allele for flower variegation in the common morning glory. Plant J. 16:371–376.

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.

Hamrick JL, Godt MJW. 1990. Allozyme diversity in plant species. In: Brown AHD, Clegg MT, Kahler AL, Weir BS, editors. Plant population genetics, breeding, and genetics resources. Sunderland (MA): Sinauer. p. 43–63.

Hudson RR. 2001. Two-locus sampling distributions and their application. Genetics. 159:1805–1817.

Hudson RR. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics. 18:337–338.

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.

Hudson RR, Kreitman M, Aguadé M. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics. 116:153–159.

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.

Ingvarsson PK. 2005. Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European aspen (*Populus tremula L., Salicaceae*). Genetics. 169:945–953.

Kado T, Yoshimaru H, Tsumura Y, Tachida H. 2003. DNA variation in a conifer, Cryptomeria japonica (*Cupressaceae sensu lato*). Genetics. 164:1547–1559.

Kobayashi M, Lin JZ, Davis J, Francis L, Clegg MT. 2000. Quantitative analysis of avocado outcrossing and yield in California using RAPD markers. Sci Hortic-Amsterdam. 86:135–149.

Krutovsky KV, Neale DB. 2005. Nucleotide diversity and linkage disequilibrium in cold-hardiness- and wood-quality-related candidate genes in Douglas fir. Genetics. 171:2029–2041.

Liu A, Burke JM. 2006. Patterns of nucleotide diversity in wild and cultivated sunflower. Genetics. 173:321–330.

Morrell PL, Toleno DM, Lundy KE, Clegg MT. 2005. Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare ssp spontaneum*) despite high rates of self-fertilization. Proc Natl Acad Sci USA. 102:2442–2447.

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.

Myers SR, Griffiths RC. 2003. Bounds on the minimum number of recombination events in a sample history. Genetics. 163:375–394.

Ness RW, Wright SI, Barrett SC. 2010. Mating-system variation, demographic history and patterns of nucleotide diversity in the Tristylous plant *Eichhornia paniculata*. Genetics. 184:381–392.

Nickerson DA, Tobe VO, Taylor SL. 1997. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucleic Acids Res. 25:2745–2751.

Nielsen R, Hubisz MJ, Clark AG. 2004. Reconstituting the frequency spectrum of ascertained single-nucleotide polymorphism data. Genetics. 168:2373–2382.

Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R, et al. 2005. The pattern of polymorphism in *Arabidopsis thaliana*. PLoS Biol. 3:e196.

Onge KR, Källman T, Slotte T, Lascoux M, Palmé AE. 2011. Contrasting demographic history and population structure in *Capsella rubella* and *Capsella grandiflora*, two closely related species with different mating systems. Mol Ecol. 20:3306–3320.

Padhukasahasram B, Marjoram P, Nordborg M. 2004. Estimating the rate of gene conversion on human chromosome 21. Am J Hum Genet. 75:386–397.

Plagnol V, Padhukasahasram B, Wall JD, Marjoram P, Nordborg M. 2006. Relative influences of crossing over and gene conversion on the pattern of linkage disequilibrium in *Arabidopsis thaliana*. Genetics. 172:2441–2448.

Rausher MD, Augustine D, Vanderkool A. 1993. Absence of pollen discounting in a genotype of *Ipomoea purpurea* exhibiting increased selfing. Evolution. 47:1688–1695.

Rausher MD, Miller RE, Tiffin P. 1999. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. Mol Biol Evol. 16:266–274.

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

Schoen DJ, Clegg MT. 1985. The influence of flower color on outcrossing rate and male reproductive success in *Ipomoea purpurea*. Evolution. 39:1242–1249.

Song YS, Ding ZH, Gusfield D, Langley CH, Wu YF. 2006. Algorithms to distinguish the role of gene-conversion from single-crossover recombination in the derivation of SNP sequences in populations. Lect Notes Comput Sc. 3909:231–245.

Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet. 68:978–989.

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.

Thornton KR. 2009. Automating approximate Bayesian computation by local linear regression. BMC Genet. 10:35.

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.

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.

Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS. 2005. The effects of artificial selection on the maize genome. Science. 308:1310–1314.

Wright SI, Charlesworth B. 2004. The HKA test revisited: a maximum-likelihood-ratio test of the standard neutral model. Genetics. 168:1071–1076.

Received April 1, 2012; Revised May 29, 2012;
Accepted July 1, 2012

Corresponding Editor: Kenneth Olsen