Heredity 86 (2001) 545–556 Received 29 August 2000, accepted 12 January 2001

Population genetic structure of the cleistogamous plant species *Viola pubescens* Aiton (Violaceae), as indicated by allozyme and ISSR molecular markers

THERESA M. CULLEY* & ANDREA D. WOLFE
Department of Evolution, Ecology & Organismal Biology, The Ohio State University, 1735 Neil Avenue, Columbus, OH U.S.A. 43210-1293

Few studies of genetic variation have focused on species that reproduce through both showy, chasmogamous (CH) flowers and self-pollinated, cleistogamous (CL) flowers. Using two different techniques, genetic variation was measured in six populations of *Viola pubescens* Aiton, a yellow-flowered violet found in the temperate forests of eastern North America. Results from eight allozyme loci showed that there was considerable genetic variation in the species, and population structuring was indicated by the presence of unique alleles and a θ (F_{ST}) value of 0.29. High genetic variation was also found using ISSR (inter-simple sequence repeat) markers, and population structuring was again evident with unique bands. *Viola pubescens* appears to have a true mixed-mating system in which selfing through CL and CH flowers contributes to population differentiation, and outcrossing through CH flowers increases genetic variation and gene flow among populations. Overall, allozyme and ISSR techniques yielded similar results, indicating that ISSR markers show potential for use in population genetic studies.

Keywords: cleistogamy, genetic variation, ISSR, population structure, selfing, *Viola pubescens*.

Introduction

The amount and distribution of genetic variation among populations can be influenced by the mating system of a species (Hamrick & Godt, 1990; Schoen *et al.*, 1996). Wright (1946, 1969) demonstrated that selfing and inbreeding may result in the loss and/or fixation of certain alleles within a population, resulting in lower genetic diversity. On a global scale, selfing may promote population substructuring as genetically isolated populations undergo genetic drift and become fixed for different alleles. In contrast, outcrossing in a population may yield high levels of genetic variation, due in part to recombination events and to occasional gene flow via pollen from neighbouring populations. Overall, populations of outcrossing species may be genetically similar to one another since they share a high number of alleles. However, some plant species have a mixed-mating system in which both outcrossing and selfing occur within a population. In these species, the same individual produces both open, chasmogamous (CH) flowers and closed, cleistogamous (CL) flowers. The elaborate CH flowers are often attractive to pollinators and are thought to facilitate outcrossing, whereas the less conspicuous CL flowers are automatically self-pollinated. These species, known as CH/CL or cleistogamous species, are widely distributed throughout the world and are present in at least 56 angiosperm families, comprising 287 species (Lord, 1981).

To date, the population genetic structure of CH/CL species has been studied in only six taxa (Schoen, 1984; Knight & Waller, 1987; Lesica *et al.*, 1988; Cole & Biesboer, 1992; Sun, 1999). Overall, these studies showed that there was usually high inbreeding, little or no genetic variability within populations, and large genetic differences among populations of CH/CL species; this is consistent with highly selfing species. In many of these taxa, selfed CL flowers were responsible for the majority of seed production (Schoen, 1984; Sun, 1999), but rates of self-pollination in CH flowers were also substantial (Clay, 1982; Cole & Biesboer, 1992; Stewart, 1994). Selfing in CH flowers could occur

*Correspondence and present address: Department of Ecology and Evolutionary Biology, University of California, Irvine, CA, U.S.A. 92697–2525. E-mail: tculley@uci.edu

© 2001 The Genetics Society of Great Britain.

through geitonogamy, pollinator-mediated selfing within the same flower, or autogamy (Lloyd & Schoen, 1992). Sun (1999) suggested that additional studies of more CH/CL species were needed to determine if other species shared this pattern of genetic structure.

We studied the population genetic structure of a CH/CL violet species, *Viola pubescens* Aiton (Violaceae), since nothing is known of its genetic variation (little is known of the genetic variation of *Viola* in general) and CH flowers appear to be more important in reproduction than in other CH/CL species. CH flowers were responsible for most of the seeds that were successfully dispersed in an Ohio population of *V. pubescens* in 1997 (22% of CH flowers successfully set seed, compared to 9% of CL flowers; T. Culley, unpublished data). The selfing rate in CH flowers is intermediate (> 0.50; T. Culley, unpublished data), and CH flowers are capable of delayed selfing if left unvisited by insect pollinators, which include carpenter bees, halictid bees, bee flies and skipper butterflies (T. Culley, pers. obs.). Gene flow via seed dispersal is probably not significant because seeds are only dispersed a few metres within isolated populations via ballistic and ant dispersal mechanisms.

Most published studies of population structure involving CH/CL species have been conducted using allozymes. However, allozymes may underestimate diversity, since additional variation has sometimes been seen when other molecular markers have been employed (Esselman *et al.*, 1999 and references therein). We were therefore interested in measuring genetic structure using allozymes and an additional molecular marker known as inter-simple sequence repeats (ISSRs). This relatively new technique is similar to that for RAPDs (Wolfe & Liston, 1998), except that the ISSR primer consists of a di- or trinucleotide simple sequence repeat with a 5′ or 3′ anchoring sequence of 1–3 nucleotides. Compared with RAPD primers, the ISSR primer sequence is usually larger, allowing for a higher primer annealing temperature which results in greater band reproducibility than RAPD markers.

The principle of ISSRs is that primer sites are dispersed throughout the genome so that there is a high chance of the primer binding to two sites located on opposing DNA strands within an amplifiable distance of one another. Thus, single-primer amplifications often result in a high degree of polymorphic bands (see Wolfe *et al.*, 1998a,b), which is useful in within-population studies when allozymes show little or no variation. Limitations of the ISSR technique are that bands are scored as dominant markers (see below) and that genetic diversity estimates are based on diallelic characters (band presence or absence). ISSRs have been used in cultivated species since 1994, but they have only recently been employed to study population variation in vascular plants (see Wolfe & Liston, 1998; Wolfe *et al.*, 1998a,b). Only one other study has directly compared genetic diversity estimates based on allozyme and ISSR data (Esselman *et al.*, 1999).

Here we report the findings of a genetic study of the CH/CL species, *Viola pubescens*. Our first objective was to determine the population genetic structure of *V. pubescens*, using six populations distributed over a geographical range. Our second goal was to compare results for allozymes with those for ISSRs.

Materials and methods

The study organism

Viola pubescens is a herbaceous perennial, commonly found in the understorey of mixed mesic forests in north-eastern North America. This stemmed species produces yellow chasmogamous (CH) flowers in the early spring (mid-April to May) and small, self-pollinated cleistogamous (CL) flowers after the forest canopy forms (May to September). The species is nonclonal, making it easy to identify genets in the field. Foliage may be either densely pubescent (var. *pubescens* Hort.) or glabrous (var. *scabriuscula* Schwein.), although seed capsules of both varieties may be hairy or glabrous (Ballard, 1994). The two varieties also differ in their numbers of stems and basal leaves, and teeth number on leaf margins (Lévesque & Dansereau, 1966; Cain, 1967; Ballard, 1994).

Molecular analyses

Genetic variation was analysed in five populations of *V. pubescens* var. *scabriuscula* and one population of *V. pubescens* var. *pubescens* located within the species range (Fig. 1). On a local scale, three populations of var. *scabriuscula* (Etter, Hill, and Stump) were situated in Crawford County, Ohio (OH) within agricultural woodlots, which were chosen because of their similar sizes (> 21 ha). These populations were compared to a population of var. *scabriuscula* at Ohio Wesleyan University's Bohannan Scientific Preserve (40.5 ha) in nearby Delaware and Morrow Counties, OH. Finally, a distant population in Emmet County, Michigan (> 50 ha; near the University of Michigan's Biological Station) contained intermingled populations of var. *scabriuscula* and var. *pubescens* individuals. These varieties were analysed separately because their genetic relationship was unknown. For simplicity, the var. *pubescens* population will be referred to by its variety type and the var. *scabriuscula* population as the 'Michigan' or 'MI' population.

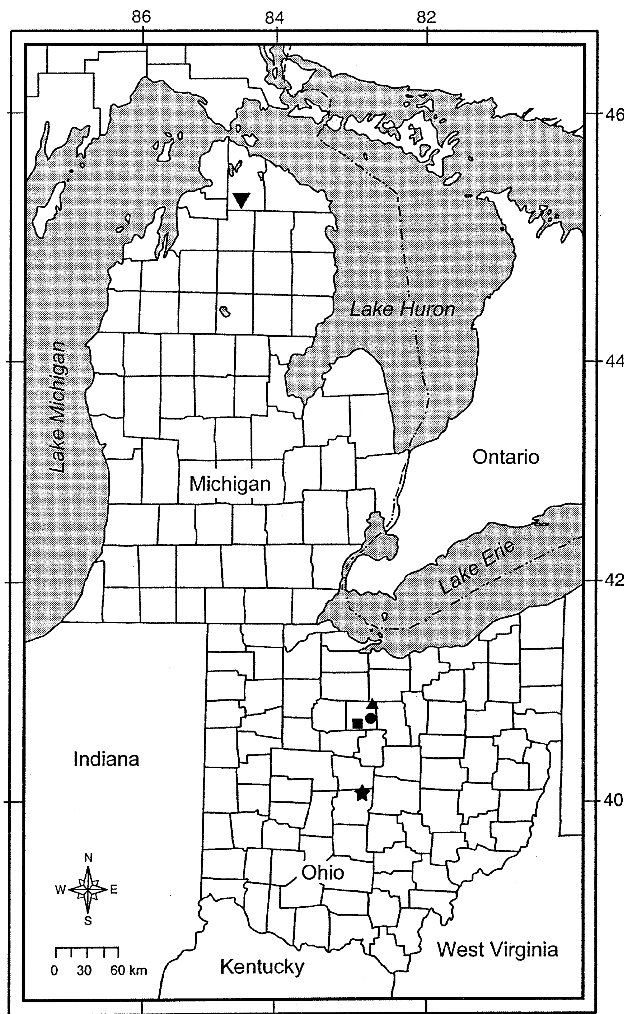

Fig. 1 Five sites of *Viola pubescens* sampled in the study: the Michigan site (▼) in Emmet County, Michigan; the Bohannan site (★) in Delaware and Morrow Counties, Ohio; the Etter (●), Hill (■), and Stump (▲) sites in Crawford County, Ohio. Numbers represent degrees latitude and longitude.

Leaf and CL bud tissue was collected from 36 to 46 plants located at least 2 m apart in each of the OH populations in spring 1997, and at the MI site in 1998; 191 plants were sampled in the study. The tissue samples were stored on ice or dry ice for transport back to the laboratory, where they were frozen in liquid nitrogen and stored at −80°C.

Allozyme analysis

Small portions of samples from all populations were ground using a *Sorghum* extraction buffer (Morden *et al.*, 1987) and the supernatant was absorbed onto filter paper wicks that were stored at −80°C. Two buffer systems were used to resolve seven enzymes, resulting in a total of nine loci and 31 alleles. Isocitrate dehydrogenase (IDH; EC 1.1.1.42), phosphogluconate dehydrogenase (PGD; EC 1.1.1.44), and shikimate dehydrogenase (SKD; EC 1.1.1.25) were resolved using a morpholine–citrate buffer system (pH 6.1; Clayton & Tretiak, 1972). Aminopeptidase (AMP; EC 3.4.11.1), malate dehydrogenase (MDH; EC 1.1.1.37), glucose-6-phosphate isomerase (GPI; EC 5.3.1.9), and phosphoglucomutase (PGM; EC 5.4.2.2) were resolved using a histidine–citrate buffer (pH 5.7; Stuber *et al.*, 1977). Gels were made of 11–12% potato starch (Starch Art, Smithville, TX, USA). Samples from all five populations were run alongside markers of known genotypes to ensure consistent scoring. Gels were run at a constant voltage for 5 h at 170–180 V (morpholine–citrate buffer) or 190–210 V (histidine–citrate buffer).

The staining recipe for AMP was from Morden *et al.* (1987), and all others were those of Wendel & Weeden (1989) with slight modifications: an agar overlay and a stain bath was used for GPI and IDH, respectively. IDH, MDH, GPI, and PGD were scored as dimers and AMP, PGM, and SKD were scored as monomers. Alleles were designated as letters representing band migration distance, with '*a*' assigned to the most anodal allozyme. Two MDH and AMP loci were used; for the other allozymes, single loci were resolved. Mendelian inheritance of these loci was confirmed using segregation patterns from CL seeds.

Measures of genetic variation within and among the populations were calculated using the GDA software package (Lewis & Zaykin, 1999). For each population, the following measurements were computed: allele frequencies per locus, number of alleles per locus (A) and per polymorphic locus (A_p), the percentage of polymorphic loci (P; 0.95 level), observed heterozygosity (H_o), and Nei's (1978) unbiased estimate of gene diversity, the expected proportion of heterozygous loci per individual (H_e). Both Hardy–Weinberg equilibrium and linkage disequilibrium between pairs of loci were tested in all populations, using sequential Bonferroni corrections of the rejection level (Rice, 1989).

To compare the amount of total genetic variation partitioned within and among populations, F-statistics and genetic distances were computed using allozyme data. Although it may be inappropriate to use the two methods on the same data set because of differences in the underlying assumptions (Reynolds *et al.*, 1983; Weir, 1996; p. 197), both were calculated for the following reasons. First, Nei's (1978) genetic distance was estimated because both drift and mutation may be important in the populations used in this study (divergence due to mutation might occur between geographically distant populations, whilst drift could be more significant in neighbouring, but isolated

populations within Crawford County). *F*-statistics were calculated for comparison with investigations of other CH/CL species.

Nei's (1978) unbiased genetic distances were computed for each pair of populations using GDA, and a neighbour-joining tree was generated in NTSYSpc (Rohlf, 1998) to visualize population relationships. *F*-statistics were calculated using the method of Weir & Cockerham (1984), which incorporates effects of small and uneven sample numbers. Weir & Cockerham's f is analogous to Wright's (1951) F_{IS}, and measures the correlation of genes within individuals in populations. θ (similar to F_{ST}) measures the amount of differentiation among populations relative to the total diversity. Upper and lower 95% confidence intervals were generated for f and θ in GDA by bootstrapping across loci (see Weir, 1996) using 1000 replicates.

ISSR analysis

DNA was extracted with a modified mini-prep technique of Doyle & Doyle (1987), using the original tissue samples that were analysed with allozyme electrophoresis. Once DNA was extracted, it was stored at −20°C until further analysis.

Three simple sequence repeats (SSR) were used as primers to generate a total of 83 bands in single-primer reactions. The trinucleotide primer, Mao [$(CTC)_4RC$], yielded 23 bands, and two dinucleotide primers, 17898 [$(CA)_6RY$] and 844 [$(CT)_8RG$] generated 32 and 28 bands, respectively. All single-tube reactions were optimized by adjusting the amounts of $MgCl_2$, Taq polymerase (GIBCO/BRL) and the primer annealing temperature. Mao was best optimized using 3 mM $MgCl_2$, 0.25 U Taq polymerase and a 45°C annealing temperature. Reactions with primer 17898 contained 3 mM $MgCl_2$ and 0.50 U Taq polymerase at 45°C, whereas primer 844 was best resolved with 2 mM $MgCl_2$ and 0.25 U Taq polymerase at 46°C. The remaining cocktail ingredients per reaction (25 μL total in each) were the same for each of the three primers: 0.5 μL DNA, 0.2 mM dNTPs, 0.4 mM primer, and 1× Taq DNA polymerase buffer. The polymerase chain reaction (PCR) was conducted in a Stratagene RoboCycler 40 and the programme was 1.5 min at 94°C; 35 × 45 s at 94°C, 45 s at 45°C or 46°C, 1.5 min at 72°C; 45 s at 94°C, 45 s at 45°C or 46°C, 5 min at 76°C; 6°C soak.

Following PCR, 1.5 μL bromophenol blue marker dye was added to each reaction and the samples were loaded onto a 1.2% agarose gel in 1× TAE buffer. Additionally, 1 kb ladders (GIBCO/BRL), and negative and positive controls were loaded onto each gel. The gels were run at constant voltage (≈144 V–150 V) until the marker dye migrated 10 cm (≈2 h). Each gel was stained with ethidium bromide and digitized under UV light using the Alpha Innotech imaging system (Alpha Innotech, San Leandro, CA, USA). The images were analysed using the BioMax 1D software package (Eastman Kodak), which assigns a fragment size to each band using an algorithm based on the 1 kb ladder. These fragment sizes were used to assign loci for each primer, and bands for each assigned locus were scored as diallelic (1 for band present; 0 for band absent). To ensure repeatability of the results, a replicate of each gel was run after repeating the PCR amplification using the previously extracted DNA. Only bands that were common to both gels were used in the final analysis and 13 individuals were removed because of non-repeatability of bands with one or more primers. This resulted in a total of 178 individuals sampled in the ISSR study.

Analysis of ISSR data is not as straightforward as allozyme data because of the dominant nature of the markers. As with RAPD or AFLP data, the presence of a band can denote either a dominant homozygote or a heterozygote, so that it is generally not possible to distinguish between genotypes. Hardy–Weinberg equilibrium is often assumed so that allele frequencies can be calculated for the genetic analyses (see Lynch & Milligan, 1994), although this assumption could be easily violated with dominant markers. Another problem is that band absence may not always indicate a homozygous recessive genotype, but may be caused by loss of a primer annealing site (because of nucleotide sequence differences), insertions or deletions in the fragment between the two primer sites, or experimental error. Consequently, the absence of a given band in two individuals may not be due to identical ancestral mutations.

To avoid these problems, genetic statistics were employed that only used the bands themselves, without invoking Hardy–Weinberg equilibrium. It was first assumed that marker alleles (bands) from different loci did not co-migrate to the same position on the gel, bands shared by two individuals were descended from a common ancestor, and each locus consisted of only two alleles that segregated in a Mendelian fashion. To characterize ISSR variation, the numbers of shared and unique bands were calculated, along with the percentages of polymorphic and fixed loci in the populations. To measure population differentiation, the Nei & Li (1979) similarity coefficient was used to compare the number of bands that were shared between individuals or populations (excluding shared absences). This coefficient was calculated for each pair of populations using !WAVSIML (V. Ford, unpublished; available via anonymous FTP at 140.254.12.151 in

incoming\ISSR; see Crawford *et al*., 1998 for formulae) and genetic distances were then computed for each pair of populations as (1-similarity). These distances were used to construct a population-level neighbour-joining tree in NTSYSpc (Rohlf, 1998). To determine how individuals from all populations clustered together, a principal coordinates analysis (PCoA) was performed in NTSYSpc using Nei and Li genetic distances generated among all pairs of individuals across the populations. This matrix was generated in !WXDNL (V. Ford, unpublished).

Results

Allozyme variation

The allozyme analysis revealed a large amount of genetic variation in *Viola pubescens*. In each population, five to eight of the nine loci examined were polymorphic, and only one locus (*Mdh-2*) was monomorphic in all populations (Table 1). At five polymorphic loci (*Idh*, *Mdh-1*, *Skd-1*, *Pgd-1* and *Pgm-1*), heterozygous individuals were detected in all populations. Since *Mdh-1* was fixed in the heterozygous condition (or was a duplicated locus) and did not behave in a Mendelian fashion, it was removed from all analyses. With the remaining eight loci, the mean observed heterozygosity was 0.35 (Table 2), which was slighter larger but not significantly different than the value expected under Hardy–Weinberg conditions ($H_e = 0.32$; *t*-test, $t_{10} = 0.96$, $P > 0.20$).

The mean number of alleles per locus (A) ranged from 2.1, within the Crawford County populations (Etter, Hill and Stump), to 3.4, in the Bohannan population (Table 2). The Crawford County and var. *pubescens* populations had a lower percentage of polymorphic loci (50%–62%) than either the Bohannan or Michigan populations (both 88%). Observed and expected heterozygosity were consistently high in all populations (>0.28), with the greatest difference occurring in the Bohannan population (Table 2). The Crawford County populations and the var. *pubescens* population had a slightly greater number of observed heterozygotes than expected, although none of these differences were significant (paired *t*-test; $t_5 = 0.78$, $P = 0.47$).

Population structuring was evident in the allozyme analysis. The Crawford County populations (Etter, Hill, and Stump) contained only a subset of the genetic variation present in the other populations, and two alleles (*Idh-1*^{f} and *Gpi-1*^{a}) were missing from the Crawford County group (Table 1). In addition, unique alleles were found in the Bohannan population (*Skd-1*^{d}, *Amp-1*^{c}, *Gpi-1*^{d}), and in the Michigan population (*Pgd-1*^{a}). The Michigan and var. *pubescens* populations also shared one unique allele (*Idh-1*^{a}). In most cases, these unique alleles were at very low frequencies and it is possible that increased sampling would have revealed their existence in other populations.

Distinct genetic differences were detected between the two varieties of *V. pubescens* at the Michigan site. Both varieties were nearly fixed for different alleles at two allozyme loci. The *Gpi-1*^{a} and *Amp-1*^{b} alleles were present at high frequencies in var. *pubescens* individuals, whilst *Gpi-1*^{b} and *Amp-1*^{a} were at high frequencies in var. *scabriuscula* individuals (Table 1). In these cases, it was almost possible to assign individuals to the correct variety solely on the basis of these allozyme genotypes. These observations were verified after re-collecting samples from 20 plants each of var. *pubescens* and var. *scabriuscula* at the same site in spring 1999. There were no significant differences from the original allele frequencies in either var. *pubescens* ($\chi^2_1 = 0.39$, $P > 0.05$) or var. *scabriuscula* ($\chi^2_1 = 0.04$, $P > 0.05$). The identity of individuals using *Gpi-1* and *Amp-1* could be correctly predicted in 94.7% of all cases.

Significant deviations from Hardy–Weinberg equilibrium occurred at only one locus (*Pgd-1*) in all populations. The remaining loci were in Hardy–Weinberg equilibrium in at least one population. The direction of deviations was not consistent across comparisons; both heterozygote excess and deficiency were observed. In addition, significant linkage disequilibrium was not detected within most populations. In the Bohannan and Michigan populations however, 3% and 14%, respectively, of the 28 pairwise comparisons between loci were significant.

F-*statistics*

Overall, the inbreeding coefficient within populations (f) calculated from the allozyme data (−0.09) was not significantly different from zero (Table 3). Negative values indicate heterozygote excess, which was largely due to high heterozygote frequencies in the Crawford County and var. *pubescens* populations. The range of f-values across loci was very large (Table 3), suggesting that evolutionary forces other than nonrandom mating may be acting differently on individual loci. However, this range is not unique among CH/CL species (e.g. Cole & Biesboer, 1992). The value of θ derived from the allozyme data (0.29) was significantly different from zero (Table 3), indicating that genetic differentiation occurred among the sampled populations.

ISSR variation

High genetic variation was also observed using ISSR markers. At the species level, 100% of the loci were polymorphic even though primers were not

Table 1 Allele frequencies for nine allozyme loci in five populations of *Viola pubescens* var. *scabriuscula* and one population of *V. pubescens* var. *pubescens* (var. *pub*). *N* is the number of individuals sampled in each population for each locus

	Population					
Locus	Etter	Hill	Stump	Bohannan	Michigan	var. *pub*
Idh-1						
(*N*)	38	35	36	31	23	23
a	0.00	0.00	0.00	0.00	0.04	0.02
b	0.07	0.06	0.06	0.14	0.11	0.76
c	0.14	0.31	0.42	0.27	0.00	0.04
d	0.67	0.51	0.25	0.45	0.54	0.06
e	0.12	0.11	0.28	0.11	0.20	0.06
f	0.00	0.00	0.00	0.02	0.11	0.04
Pgd-1						
(*N*)	38	35	34	33	23	22
a	0.00	0.00	0.00	0.00	0.04	0.00
b	0.18	0.03	0.03	0.14	0.04	0.00
c	0.12	0.41	0.37	0.20	0.46	0.50
d	0.53	0.23	0.28	0.46	0.04	0.00
e	0.17	0.33	0.32	0.21	0.41	0.50
Skd-1						
(*N*)	38	35	35	32	23	23
a	0.00	0.04	0.00	0.02	0.41	0.20
b	0.50	0.13	0.13	0.28	0.04	0.20
c	0.50	0.83	0.87	0.58	0.54	0.60
d	0.00	0.00	0.00	0.12	0.00	0.00
Pgm-1						
(*N*)	37	35	36	31	21	22
a	0.00	0.00	0.03	0.02	0.07	0.00
b	0.55	0.53	0.64	0.94	0.81	0.96
c	0.45	0.47	0.33	0.05	0.12	0.04
Mdh-1						
(*N*)	38	35	36	35	23	23
a	0.50	0.50	0.50	0.50	0.50	0.50
b	0.50	0.50	0.50	0.50	0.50	0.50
Mdh-2						
(*N*)	38	35	36	35	23	23
a	1.00	1.00	1.00	1.00	1.00	1.00
Amp-1						
(*N*)	38	35	34	32	23	22
a	1.00	0.96	0.84	0.80	0.83	0.04
b	0.00	0.04	0.16	0.11	0.17	0.96
c	0.00	0.00	0.00	0.09	0.00	0.00
Amp-2						
(*N*)	38	35	36	32	21	20
a	0.00	0.01	0.00	0.03	0.10	0.00
b	1.00	0.97	0.97	0.39	0.00	0.30
c	0.00	0.01	0.03	0.58	0.90	0.70
Gpi-1						
(*N*)	38	35	36	36	23	23
a	0.00	0.00	0.00	0.03	0.20	0.87
b	0.99	1.00	1.00	0.83	0.80	0.13
c	0.01	0.00	0.00	0.12	0.00	0.00
d	0.00	0.00	0.00	0.01	0.00	0.00

Table 2 Genetic variability at eight allozyme loci in five populations of *Viola pubescens* var. *scabriuscula* and one population of *V. pubescens* var. *pubescens* (var. *pub*). Mean sample sizes per locus (N), mean number of alleles per locus (A) and per polymorphic locus (A_p), percentage of polymorphic loci (P), Nei's (1978) unbiased estimate of the expected proportion of heterozygous loci per individual (H_e), and observed heterozygosity (H_o) are shown. A locus was considered polymorphic if the frequency of the most common allele was less than 0.95. Standard errors are shown in parentheses

Population	N	A	A_p	P	H_e	H_o
Etter	37.9	2.1	3.0	50.0	0.28 (0.10)	0.39 (0.15)
Hill	35.0	2.5	3.2	50.0	0.28 (0.10)	0.37 (0.15)
Stump	35.4	2.4	3.0	62.5	0.30 (0.10)	0.32 (0.12)
Bohannan	32.8	3.4	3.7	87.5	0.41 (0.09)	0.30 (0.10)
Michigan	22.5	2.9	3.1	87.5	0.37 (0.08)	0.36 (0.16)
var. *pub*	22.2	2.5	3.0	62.5	0.29 (0.08)	0.35 (0.14)
Total	31.0	2.6	3.2	66.7	0.32 (0.02)	0.35 (0.01)

Table 3 Summary of F-statistics (Weir & Cockerham, 1984) at eight allozyme loci analysed in six populations of *Viola pubescens*. The statistic f measures the correlation of genes within individuals in populations, and θ measures the amount of differentiation among populations relative to the total diversity. Asterisks indicate a monomorphic locus, and 95% confidence intervals (CI) were derived from bootstrapping across loci with 1000 replicates

Locus	f	θ
Idh-1	0.08	0.18
Pgd-1	−0.25	0.10
Skd-1	−0.44	0.15
Pgm-1	−0.47	0.16
Mdh-2	***	***
Amp-1	0.66	0.52
Amp-2	0.12	0.63
Gpi-1	0.80	0.59
Mean	−0.09	0.29
Upper CI	0.21	0.48
Lower CI	−0.31	0.16

intentionally selected for high variability. Within each population, over 71% of the 83 loci were polymorphic, with the Bohannan population exhibiting the highest number of polymorphic loci (84.3%). The mean proportion of polymorphic loci in the populations sampled was 77.1%.

Population structuring was evident with the ISSR data. One unique band was detected in each of the Etter and Hill populations, whilst the Bohannan and var. *pubescens* populations had two and three unique bands, respectively (data available upon request). Of the 83 loci, 36 (43%) had bands that were found in at least two individuals of all six populations. One band was fixed in the Crawford County populations and was near fixation (average frequency of 82.4%) in the remaining populations. A second band was also fixed throughout the Etter, Hill, and var. *pubescens* populations.

Genetic distance

Nei's (1978) genetic distance based on allozyme allele frequencies was estimated for all 15 pairwise comparisons between the populations (Table 4). The mean distance for all comparisons was 0.268, ranging from 0.010 to 0.760. The lowest genetic distances were found among the Crawford County populations (Etter, Hill and Stump), whereas the var. *pubescens* population was the most dissimilar to all other populations (Table 4).

Genetic distances for all pairwise comparisons were also calculated from the ISSR data using the Nei & Li (1979) distance coefficient. Overall, the ISSR pairwise distances were higher than those generated with allozymes (Table 4). As with the allozyme data, the Crawford County populations were most genetically similar to each other. The var. *pubescens* population was again most dissimilar to other populations, although the Michigan population also showed large genetic distances in some comparisons (Table 4). The neighbour-joining phenogram derived from allozyme data showed the same topology as the ISSR-based phenogram, although there were slight differences in branch lengths (Fig. 2).

Genetic distances were also examined between individuals within the total population, using principal coordinates analysis (PCoA; Fig. 3). Samples from the Crawford County populations mostly grouped together, and samples from the Michigan and var. *pubescens* populations (located at the same site) sorted into a loose cluster. Most samples from the Bohannan population clustered largely with the Crawford County group, although several samples overlapped with the MI populations. As expected, var. *pubescens* samples clustered furthest from the other populations, all of which consisted of a different species variety. This plot

Population	Etter	Hill	Stump	Bohannan	Michigan	var. *pub*
Etter	****	0.333	0.400	0.524	0.585	0.652
Hill	0.042	****	0.388	0.518	0.601	0.624
Stump	0.066	0.010	****	0.537	0.603	0.627
Bohannan	0.114	0.121	0.108	****	0.640	0.672
Michigan	0.310	0.244	0.257	0.090	****	0.650
var. *pub*	0.760	0.627	0.546	0.402	0.320	****

Table 4 Genetic distances between pairs of *Viola pubescens* populations using ISSR and allozyme data. Above the diagonal are Nei & Li's (1979) distances using 83 ISSR loci, and below the diagonal are Nei's (1978) unbiased genetic distances based on eight allozyme loci (excluding *Mdh-1*)

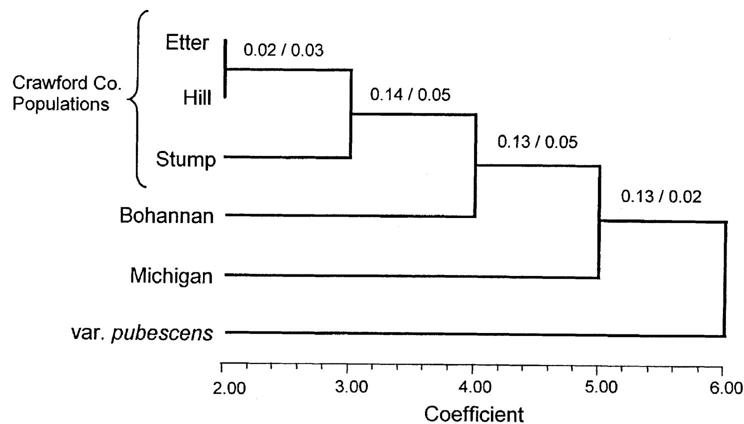

Fig. 2 Neighbour-joining phenogram of five populations of *Viola pubescens* var. *scabriuscula* and one population of var. *pubescens*. The same phenogram was obtained with allozyme data using Nei's (1978) unbiased genetic distance, and with ISSR data using the Nei & Li (1979) distance coefficient. Numbers above each line represent branch lengths (allozyme data/ISSR data).

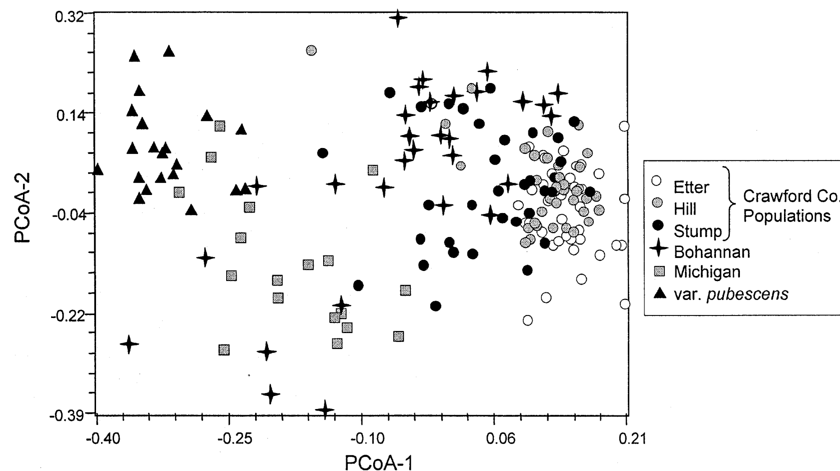

Fig. 3 Principal coordinates analysis (PCoA) using the Nei & Li (1979) distance coefficient computed from ISSR data collected from 178 individuals. All populations consisted of var. *scabriuscula* except for one population of var. *pubescens*. The proportion of total variance comprising each axis was 13.4% for axis 1 and 7.0% for axis 2.

indicates that the differentiation between the two varieties in Michigan is about as extreme as the differentiation amongst all Ohio populations.

Discussion

Population genetic structure

Our study revealed patterns of genetic diversity in *Viola pubescens* that are consistent with its mixed-mating system, yet different from other CH/CL species that appear to be highly selfed (Table 5). Sampled populations of *V. pubescens* displayed a surprisingly large amount of genetic variation with both allozyme and ISSR molecular markers. Such high levels of genetic variation are more typical of an outcrossing species (Hamrick & Godt, 1990) and may represent the effect of outcrossing through CH flowers. On the other hand, significant population substructuring ($\theta > 0$) and the presence of some unique alleles suggest that some selfing, either through CL or self-pollinated CH flowers, takes place in the populations.

Table 5 Genetic variability measures derived from allozymes for CH/CL species. In most cases, the percentage of polymorphic loci (P) was not specified as 95% or 99%. Asterisks denote measures for only polymorphic loci. See text for details of notation

Species (Family)	No. of pops	No. of loci	A	A_P	P	H_E	H_O	$F_{IS(f)}$	$F_{ST(\theta)}$	Citation
Howellia aquatilis (Campanulaceae)	4	18	1.00	—	0.00	0.00	0.00	—	—	Lesica *et al*. (1988)
Impatiens capensis (Balsaminaceae)	11	31	—	2.02	9.97	—	—	0.57*	0.46*	Knight & Waller (1987)
Lespedeza capitata (Fabaceae)	12	34	1.08	—	7.84	0.02	0.01	0.56	0.51	Cole & Biesboer (1992)
Lespedeza leptostachya (Fabaceae)	12	32	1.00	—	0.00	0.00	0.00	—	0.00	Cole & Biesboer (1992)
Microlaena polynoda (Poaceae)	1	9	1.00	—	0.00	0.00	0.00	—	—	Schoen (1984)
Scutellaria indica (Lamiaceae)	20	30	1.02	2.03	2.36	0.01	0.001	0.74*	0.92*	Sun (1999)
Viola pubescens (Violaceae)	6	8	2.63	3.18	66.67	0.32	0.35	−0.09	0.29	Present study

In general, *V. pubescens* possesses more genetic variation than most *Viola* species studied thus far. There was no allozyme variation within the Asian *V. albida* complex (Kim *et al*., 1991), and in *V. collina* (Marcussen & Borgen, 2000). Most populations of *V. suavis* and associated species were generally monomorphic (Marcussen & Nordal, 1998), and low levels of variation were found in *Viola rupestris* (Nordal & Jonsell, 1998) and in nine *Viola* taxa (Marcussen & Borgen, 2000). However, low genetic variability in the last study may reflect past glaciation events, because variability declined in a northerly direction from Mediterranean populations to those in Norway (T. Marcussen, pers. comm.). In a large Mediterranean population, three out of five allozyme systems were polymorphic (T. Marcussen, pers. comm.), which is comparable to the variation found in some populations of *V. pubescens*.

The distribution of the total genetic variation among populations in *V. pubescens* was lower than in other CH/CL species, as θ in the present study was nearly half the reported values for other species (Table 5). Enhanced gene flow via seeds in *V. pubescens* (compared to other species) is unlikely to explain this difference because seed dispersal (either ballistic or ant-facilitated) is probably not enough to transverse the wide distances of agricultural monocultures separating populations in Crawford County. A more likely reason why population structure is so different is that CH flowers are responsible for more reproduction in *V. pubescens* than in other species. Outcrossing through CH flowers could introduce more genetic variation into populations and thus prevent population differentiation. Outcrossing is also suggested by negative *f*-values (Table 3), which indicate an excess of heterozygotes. However, some selfing does occur in CH flowers, because they have an intermediate selfing rate and can self-pollinate in the absence of pollinators via a delayed-selfing mechanism (T. Culley, unpublished data).

In other CH/CL species, it appears that either CL flowers are primarily responsible for reproduction (Schoen, 1984; Sun, 1999), or that selfing rates in CH flowers are quite substantial. For example, CH flowers appear to be self-pollinated prior to flower opening in *Howellia aquatilis*, although the possibility of an occasional outcrossing event cannot be discounted (Lesica *et al*., 1988). CH flowers in some populations of *Impatiens capensis* may be self-pollinated via geitonogamy or may undergo biparental inbreeding (Waller & Knight, 1989). In some CH/CL species, selfing rates are thought to be so high that there is no genetic variation in the populations (Schoen, 1984; Lesica *et al*., 1988; Cole & Biesboer, 1992).

One interesting result of this study was the genetic differences detected between the two varieties of *V. pubescens* at the MI site. In addition to the allozyme differences noted earlier, phenograms constructed from allozyme and ISSR data (Fig. 2) show that the var. *pubescens* population is genetically dissimilar to the Michigan population (and to all other var. *scabriuscula* populations). Clustering differences were also seen in the principal coordinates analysis (Fig. 3), although there was some slight overlap of the two varieties. These two varieties were initially considered separate species because of several key differences (see above; Lévesque & Dansereau, 1966; Cain, 1967), but were grouped

together because of several morphologically intermediate herbarium specimens (Ballard, 1994). It would be interesting to know if genetic differences are found in additional populations, and if the two varieties are cross-compatible.

Comparison of allozymes and ISSRs

Overall, the allozyme and ISSR techniques gave similar results in this study, with both detecting a large amount of genetic variation (e.g. $P > 0.50$). Even though genetic distances were computed differently for allozyme and ISSR data, the relationships among pairwise genetic distances were similar for both techniques. In addition, the topology of neighbour-joining phenograms constructed from both data sets was identical, although the branch lengths differed slightly.

Presently, only one other study (Esselman *et al*., 1999) has compared allozymes with ISSR markers. These authors detected very little variation with allozymes in the clonal plant species, *Calamagrostis porteri* ssp. *insperata*, but over 10% of ISSR bands were variable. In fact, many studies of natural populations that have compared allozymes with other DNA techniques, such as RAPDs, have found that allozymes sometimes express lower amounts of genetic variation (e.g. Sun *et al*., 1999; Sun, 1999).

The similarity of the overall conclusions based on allozymes and ISSR markers in this investigation shows the potential of ISSR markers for population genetic studies. Although dominant markers do not provide the same accuracy of estimation as codominant markers (see Lynch & Milligan, 1994), ISSR markers may provide a suitable alternative to allozymes, especially in cases involving rare or clonal species that typically express low levels of allozyme variation. Compared with widely used RAPD markers, ISSRs are advantageous because they have higher band repeatability. As this is only the second population genetic study to compare allozyme and ISSR markers, additional investigations are needed in other species to create a benchmark with which to compare future results.

One caveat for population genetics studies involving ISSRs or other dominant markers is that the method of analysis must be selected carefully, due to inherent limitations, such as assumptions of Hardy–Weinberg equilibrium and random mating (Lynch & Milligan, 1994). Potential methods include the analysis of molecular variance (AMOVA) which generates ϕ_{st} (not yet directly comparable with θ generated from codominant markers; Stewart & Excoffier, 1996), the Shannon–Weaver diversity statistic, which uses band phenotypes and band absences (Whitkus *et al*., 1998), and the Jaccard and the Nei and Li coefficients, which also use band phenotypes but exclude shared band absences. Whichever method is chosen, the results must be carefully interpreted in light of the assumptions of each technique.

Conclusions

Viola pubescens shows great genetic variation within populations and moderate levels of differentiation among populations, for both allozyme and ISSR markers. This genetic structure is very different from other CH/CL species studied thus far, perhaps because outcross-pollinated CH flowers play a more important role in *V. pubescens* than in other species. Additional studies measuring genetic variation in other species would be especially helpful to determine if the high levels of genetic variation detected in *V. pubescens* are unique among other CH/CL species.

Acknowledgements

We wish to thank R. Culley, P. Doherty, T. Grubb, Jr., and A. Snow for field assistance, H. Ballard, S. Datwyler, T. Marcussen, and C. Randle for technical advice, and D. Crawford and two anonymous referees for comments on the manuscript. Permission to collect samples was given by the Ohio Wesleyan University and the University of Michigan's Biological Station. This project was supported by a 1998 Karling Graduate Student Research Award from the Botanical Society of America.

References

BALLARD, H. E. 1994. Violets of Michigan. *Michigan Bot*., **33**, 131–199.

CAIN, S. A. 1967. Studies on the stemmed yellow violets of eastern North America. II. Mass-collections of *Viola pubescens* and *V. eriocarpa* in the Michigan area. *Naturaliste Canadien*, **94**, 79–129.

CLAY, K. 1982. Environmental and genetic determinants of cleistogamy in a natural population of the grass *Danthonia spicata*. *Evolution*, **36**, 734–741.

CLAYTON, J. W. AND TRETIAK, D. N. 1972. Amine-citrate buffers for pH control in starch gel electrophoresis. *J. Fish. Res. Board Can*., **29**, 1169–1172.

COLE, C. T. AND BIESBOER, D. D. 1992. Monomorphism, reduced gene flow, and cleistogamy in rare and common species of *Lespedeza* (Fabaceae). *Am. J. Bot*., **79**, 567–575.

CRAWFORD, D. J., ESSELMAN, E. J., WINDUS, J. L. AND PABIN, C. S. 1998. Genetic variation in running buffalo clover (*Trifolium stoloniferum*: Fabaceae) using random amplified polymorphic DNA markers (RAPDs). *Ann. Mo. Bot. Gard*., **85**, 81–89.

DOYLE, J. J. AND DOYLE, J. L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull*., **19**, 11–15.

ESSELMAN, E. J., JIANQIANG, L., CRAWFORD, D. J., WINDUS, J. L. *ET AL*. 1999. Clonal diversity in the rare *Calamagrostis porteri* ssp. *insperata* (Poaceae): comparative results for allozymes and random amplified polymorphic DNA (RAPD) and inter-simple sequence repeat (ISSR) markers. *Mol. Ecol.*, **8**, 443–451.

HAMRICK, J. L. AND GODT, M. J. W. 1990. Allozyme diversity in plant species. In: Brown, A. H. D., Clegg, M. T., Kahler, A. L. and Weir, B. S. (eds) *Plant Population Genetics, Breeding, and Genetic Resources*, pp. 43–63. Sinauer Associates, Sunderland, MA.

KIM, K. S., SUN, B. Y., WHANG, S. S. AND CHUNG, G. H. 1991. Biosystematic study of the genus *Viola* in Korea: comparative morphology of the *Viola albida* complex. *Korean J. Bot.*, **34**, 229–238.

KNIGHT, S. E. AND WALLER, D. M. 1987. Genetic consequences of outcrossing in the cleistogamous annual, *Impatiens capensis*. I. Population-genetic structure. *Evolution*, **41**, 969–978.

LESICA, P., LEARY, R. F., ALLENDORF, F. W. AND BILDERBACK, D. E. 1988. Lack of genic diversity within and among populations of an endangered plant, *Howellia aquatilis*. *Conserv. Biol.*, **2**, 275–282.

LÉVESQUE, F. L. AND DANSEREAU, P. 1966. Études sur les violettes jaunes caulescentes de l'est de l'amérique du nord. I. Taxonomie, nomenclature, synonymie et bibliographie. *Naturaliste Canadien*, **93**, 489–569.

LEWIS, P. O. AND ZAYKIN, D. 1999. *Genetic Data Analysis: Computer program for the analysis of allelic data*, version 1.0 (d12). Free program distributed by the authors over the Internet from the GDA Home Page at http://chee.unm.edu/gda/.

LLOYD, D. G. AND SCHOEN, D. J. 1992. Self- and cross-fertilization in plants: I. Functional dimensions. *Int. J. Pl. Sci.*, **153**, 358–369.

LORD, E. M. 1981. Cleistogamy: a tool for the study of floral morphogenesis, function and evolution. *Bot. Rev.*, **47**, 421–449.

LYNCH, M. AND MILLIGAN, B. G. 1994. Analysis of population genetic structure with RAPD markers. *Mol. Ecol.*, **3**, 91–99.

MARCUSSEN, T., BORGEN, L. 2000. Allozymic variation and relationships within *Viola* subsection *Viola* (Violaceae). *Plant Syst. Evol.*, **223**, 29–57.

MARCUSSEN, T. AND NORDAL, I. 1998. *Viola suavis*, a new species in the Nordic flora, with analyses of the relation to other species in the subsection *Viola* (Violaceae). *Nordic J. Bot.*, **18**, 221–237.

MORDEN, C. W., DOEBLEY, J. AND SCHERTZ, K. F. 1987. A manual of techniques for starch gel electrophoresis of *Sorghum* isozymes. Texas Agricultural Experimental Station Miscellaneous Publication no. 1635. The Texas A & M University, College Station, TX.

NEI, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

NEI, M. AND LI, W.-H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U.S.A.*, **76**, 5269–5273.

NORDAL, I. AND JONSELL, B. 1998. A phylogeographic analysis of *Viola rupestris*: three post-glacial immigration routes into the Nordic area? *Bot. J. Linn. Soc.*, **128**, 105–122.

REYNOLDS, J., WEIR, B. S. AND COCKERHAM, C. C. 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics*, **105**, 767–779.

RICE, W. R. 1989. Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.

ROHLF, F. J. 1998. NTSYSpc: *numerical taxonomy and multivariate analysis system*, version 2.02h. Exeter Software, Setauket, NY.

SCHOEN, D. J. 1984. Cleistogamy in *Microlaena polynoda* (Gramineae): an examination of some model predictions. *Am. J. Bot.*, **71**, 711–719.

SCHOEN, D. J., MORGAN, M. T. AND BATAILLON, T. 1996. How does self-pollination evolve? Inferences from floral ecology and molecular genetic variation. *Phil. Trans. R. Soc. Ser. B*, **351**, 1281–1290.

STEWART, S. C. 1994. Genetic constraints on mating system evolution in the cleistogamous annual *Impatiens pallida*: inbreeding in chasmogamous flowers. *Heredity*, **73**, 265–274.

STEWART, C. N. AND EXCOFFIER, L. 1996. Assessing population genetic structure and variability with RAPD data: application to *Vaccinium macrocarpon* (American Cranberry). *J. Evol. Biol.*, **8**, 153–171.

STUBER, C. W., GOODMAN, M. M. AND JOHNSON, F. M. 1977. Genetic control and racial variation of β-glucosidase isozymes in maize (*Zea mays* L.). *Biochem. Genet.*, **15**, 383–394.

SUN, M. 1999. Cleistogamy in *Scutellaria indica* (Labiatae): effective mating system and population genetic structure. *Mol. Ecol.*, **8**, 1285–1295.

SUN, G.-L., DIAZ, O., SALOMON, B. AND VON BOTHMER, R. 1999. Genetic diversity in *Elymus caninus* as revealed by isozyme, RAPD, and microsatellite markers. *Genome*, **42**, 420–431.

WALLER, D. M. AND KNIGHT, S. E. 1989. Genetic consequences of outcrossing in the cleistogamous annual *Impatiens capensis*. II. Outcrossing rates and genotypic correlations. *Evolution* **43**, 860–869.

WEIR, B. S. 1996. *Genetic Data Analysis II*. Sinauer, Associates, Sunderland, MA.

WEIR, B. S. AND COCKERHAM, C. C. 1984. Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

WENDEL, J. F. AND WEEDEN, N. F. 1989. Visualization and interpretation of plant isozymes. In: Soltis, D. E. and Soltis, P. S. (eds) *Isozymes in Plant Biology*, pp. 5–45. Dioscorides Press, Portland, OR.

WHITKUS. R., DE LA CRUZ, M., MOTA-BRAVO, L. AND GÓMEZ-POMPA, A. 1998. Genetic diversity and relationships of cacao (*Theobroma cacao* L.) in southern Mexico. *Theor. Appl. Genet.*, **96**, 621–627.

WOLFE, A. D. AND LISTON, A. 1998. Contributions of PCR-based methods to plant systematics and evolutionary biology. In: Soltis, D. E., Soltis, P. S. and Doyle, J. J. (eds) *Plant Molecular Systematics II*, pp. 43–86. Kluwer Academic Publishers, Dordrecht.

WOLFE, A. D., XIANG, Q.-Y. AND KEPHART, S. R. 1998a. Diploid hybrid speciation in *Penstemon* (Scrophulariaceae). *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 5112–5115.

WOLFE, A. D., XIANG, Q.-Y. AND KEPHART, S. R. 1998b. Assessing hybridization in natural populations of *Penstemon* (Scrophulariaceae) using hypervariable inter-simple sequence repeat (ISSR) bands. *Mol. Ecol.*, **7**, 1107–1125.

WRIGHT, S. 1946. Isolation by distance under diverse systems of mating. *Genetics*, **31**, 39–59.

WRIGHT, S. 1951. The genetical structure of populations. *Ann. Eugenics*, **15**, 323–354.

WRIGHT, S. 1969. *Evolution and the genetics of populations. II. The Theory of Gene Frequencies*. University of Chicago Press, Chicago, IL.

© The Genetics Society of Great Britain, *Heredity*, **86**, 545–556.