Genetic Variation and Population Genetic Structure in *Trifolium pratense*

**M. J. Hagen and J. L. Hamrick**

*Trifolium pratense* (red clover) is a short-lived herbaceous plant native to southeastern Europe and Asia Minor. Widely used in agriculture, *T. pratense* is cultivated as an annual, winter annual, or biennial. It blooms from mid-spring to early fall and is insect pollinated and self-incompatible. Seeds are mammal and bird dispersed. Naturalized populations of *T. pratense* occur along roadsides and in old fields as well as native grasslands. Allelozyme diversity and population genetic structure were determined for nine populations of *T. pratense*. Results from 13 allelozyme loci indicate that genetic diversity is higher and population divergence is lower than expected based on the life-history characteristics of the species. We conclude that the high levels of genetic diversity found within populations of *T. pratense* suggest that these are not newly established founder populations, and that the low levels of genetic divergence seen among populations are probably due to high rates of gene flow among populations as a result of seed and pollen movement.

Agricultural breeding practices and life-history characteristics of cultivated species effect the levels of genetic diversity maintained within species and their populations. At the species level, high genetic diversity is desirable; thus germplasm collections and seed banks have been established to conserve and maintain the genetic diversity of cultivated species and their relatives (Blixt 1988; Marshall 1989; Marshall and Brown 1975). Genetic diversity within populations of crop species is often reduced, however, when the characteristics of high yield and phenotypic uniformity are selected.

Naturalized populations of cultivated species (i.e., escapes) are ultimately a product of both their biological characteristics and historical cultivation practices. Newly established naturalized populations should have low levels of genetic diversity for several reasons. First, genetic variation in the source population (i.e., a cultivated field) may be limited because the field was planted with a single strain or cultivar. Second, populations of colonizing species are often founded by a small sample of a larger population (i.e., founder effect) and as a result may exhibit reduced genetic diversity (Barrett and Shore 1989; Chakraborty and Nei 1977; Nei et al. 1975; Nei and Tajima 1981). Finally, because the competitive ability of cultivated species is often reduced in natural habitats, the number of successful founders may be small (i.e., founder selection), further reducing genetic variation sampled from the source population. These factors should contribute to low genetic diversity within populations and higher divergence among populations. In contrast, the mixing of strains due to seed movement and pollen flow between established populations should increase intrapopulation diversity and reduce differentiation among populations.

In this study we examined genetic variation within and among naturalized populations of *Trifolium pratense* L. (red clover). A short-lived herbaceous plant, *T. pratense* is cultivated as an annual, winter annual, or biennial. It is widely used in agriculture and is of considerable economic importance. Originating in southeastern Europe and Asia Minor, *T. pratense* was first cultivated in northern Europe around 1650 (Merkenschlager 1934) and was introduced to North America by European colonists (Fergus and Hollowell 1960; Piepers and Hollowell 1937). *T. pratense* is an insect-pollinated, self-incompatible diploid (2n = 14) that blooms from mid-spring to early fall (Smith et al. 1985). Seeds are dispersed by horses, cattle, humans and birds (Ridley 1930). *T. pratense* is commonly found along roadsides and in old fields, but it also invades natural and cultivated grasslands.

We compared the genetic diversity and population genetic structure in naturalized populations of *T. pratense* with plant species having similar life-history characteristics. We concluded that the high levels of genetic diversity found within populations of *T. pratense* suggest that these are not newly established founder populations, and that the low levels of genetic divergence seen among populations is probably due to high rates of gene flow among populations as a result of seed and pollen movement.

**Materials and Methods**

Nine natural populations of *T. pratense* were collected, three populations in the southeastern United States in 1992 (Jackson, North Carolina = JNC1; Clayton, Georgia = BMT1; Athens, Georgia = CCR1) and in 1993 (Jackson, North Carolina = JNC2; Clayton, Georgia = BMT2; Athens, Georgia = CCR2) and an additional three populations in 1993 from the northeastern United States (Biddeford, Maine = UNE; Kennebunk, Maine = KBS; Woodstock, Virginia = WHS) (Figure 1). Data from the 1992 and 1993 southeastern populations were pooled for analysis unless stated otherwise. From each population leaf samples were collected from 48 individual plants. Samples were stored on ice after collection and transported to the laboratory for electrophoretic analysis. Each leaf sample was cut finely, frozen with liquid nitrogen, and crushed with a mortar and pestle. Enzymes were stabilized and extracted by the addition of a potassium phosphate buffer (Mitton et al. 1979). The enzyme extracts were absorbed onto filter paper wicks and stored at −70°C.
Elephantiasis was performed on 10% starch gels according to the gel and electrode buffer systems of Soltis et al. (1983). Twelve enzyme stains resolved 13 loci. The enzymes stained were triose phosphate isomerase (TPI), alcohol dehydrogenase (ADH), glucose-6-phosphate dehydrogenase (GPDH), peroxidase (PER), malate dehydrogenase (MDH), glucosephosphatase isomerase (GPI), diaphorase (DIA), 6-phosphogluconate dehydrogenase (6-PGDH), malic enzyme (ME), β-galactosidase (β-GAL), leucine aminopeptidase (LAP), and fructose biphosphatase (FBP). Stain recipes were taken from Soltis et al. (1983), except for DIA and β-GAL, which were taken from Cheliak and Pitel (1984), and FBP, which is from Wendel and Weeden (1989). The genetic basis of the observed allozyme banding patterns was inferred from segregation patterns with reference to typical subunit structure (Gottlieb 1981; Harris and Hopkinson 1976).

Allozyme diversity was calculated for the species as a whole and on a population basis using four standard measures: percent polymorphic loci (P), mean number of alleles per polymorphic locus (AP), effective number of alleles per locus (Ae), and gene diversity (Hd). Subscripts refer to species (s) or population (p) level parameters. Percent polymorphic loci (Pp) was calculated by dividing the number of loci polymorphic in at least one population by the total number of loci analyzed. The mean number of alleles per polymorphic locus (APp) was determined by summing all the alleles observed at polymorphic loci and dividing by the total number of loci. The effective number of alleles was calculated for each locus by 1/Σp2, where p is the mean frequency of the ith allele. These values were then averaged across loci to obtain Ap. Gene diversity at each locus was calculated as \( h = 1 - \sum p_i^2 \). The mean genetic diversity (HD) is the arithmetic average of h values across all loci. At the population level, the values of PP, APp, Ap, and Hp were calculated using the equations of Hedrick (1985). Genetic parameters at the population level represent population means, whereas at the species level, parameters reflect overall genetic diversity within the species.

Deviations from Hardy–Weinberg expectations were examined for each population by calculating Wright’s fixation index (Wright 1922) and testing for significant deviations from the expected value (\( F = 0 \)) by chi-square tests (Li and Horvitz 1953).

Chi-square analyses were also performed to test for heterogeneity in allele frequencies among populations (Workman and Niswander 1970) and among years. Total genetic diversity was partitioned into components using Nei’s (1973, 1977) gene diversity statistics. For each polymorphic locus, total gene diversity (\( H_T \)) was partitioned into diversity among regions (\( D_{reg} \)), among populations within regions (\( D_{pop} \)), and within populations (\( H_{pop} \)) as follows: \( H_T = D_{reg} + D_{pop} + H_{pop} \). A temporal component to gene diversity, \( D_{year} \), was examined for the southern populations only; where total gene diversity in the southern sample (\( H_{s}\)) as partitioned into diversity among years (\( D_{year} \)), among populations within years (\( D_{pop(year)} \)), and within populations (\( H_{pop(year)} \)) as follows: \( H_{s} = D_{year} + D_{pop(year)} + H_{pop(year)} \). A measure of differentiation among populations, relative to total diversity, \( G_s \), as calculated at each locus and averaged over all polymorphic loci to obtain a species value for population divergence. Population divergence was also examined by calculating Nei’s genetic distance and identity parameters (Nei 1972) for all pairs of populations. An indirect estimate of gene flow among populations, \( N_m \), was estimated from the genetic data as

\[
N_m = (1 - G_s)/(\alpha 4G_s),
\]

where \( \alpha = [n/(n - 1)] \) and n = number of populations sampled (Crow and Aoki 1984).

Results

Ten of the 13 loci (76.9%) were polymorphic in at least one population. β-GAL, 6-PGDH, and ADH were monomorphic enzyme systems. The percent polymorphic loci within populations ranged from 61.5% in population CCR to 76.9% in population BMT, with an overall mean of 68.4%. The number of alleles per polymorphic locus was 2.60 at the species level and averaged 2.52 within populations. Expected heterozygosity for the species was \( H_s = 0.285 \) and at the population level was \( H_{pop} = 0.250 \) (Table 1).

Hardy–Weinberg expectations were observed in all but 5 of the 117 chi-square tests. Since we would expect to see 5.85
significant deviations at the $P = .05$ significance level by chance alone, we can conclude that the populations sampled are in Hardy–Weinberg equilibrium. Furthermore, mean observed heterozygosity within populations, $H_{obs} = 0.228$, was close to expectation, $H_{exp} = 0.250$, and the mean $F_{is}$ value, a measure of the deviation of random mating within populations, was not significantly different from zero ($P < .05$). These results are consistent with the predominantly outcrossing mode of reproduction known for this species (Smith et al. 1985).

Allele frequencies were significantly different among populations for 9 of the 10 polymorphic loci ($P < .001$), the exception being the GDH1 locus which had very little genetic variation. The mean proportion of total genetic variation due to differences among regions, $G_{reg}$, and among populations within regions, $G_{pop}$, was 0.013 and 0.049, respectively, indicating that most of the genetic diversity occurs within populations ($P < .05$). Allele frequencies among pairs of populations was relatively uniform, with a wide geographic distribution have an $H_{exp}$ of 0.124, a $P_{s}$ of 0.468, and an $AP_{s}$ of 2.4. The values for $T. pratense$ are generally higher than each of these parameters. Because high levels of genetic diversity are often deliberately maintained in crop species (Hamrick and Godt 1989; Marshall 1989; Muona 1989; Weber et al.

### Discussion

Genetic diversity in $T. pratense$ is high in comparison to that of most plant species. Mean genetic diversity ($H_{exp}$) was 0.285, mean percent polymorphic loci ($P_{s}$) was 76.9, and mean number of alleles per polymorphic locus ($AP_{s}$) was 2.60. These same parameters, averaged over many plant species, are $H_{exp} = 0.150$, $P_{s} = 50.5$, and $AP_{s} = 2.90$ (unpublished results based on data from Hamrick and Godt 1989). A similar pattern was observed at the population level. Mean genetic diversity within $T. pratense$ populations ($H_{pop}$) was 0.250, mean percent polymorphic loci ($P_{pop}$) was 68.3, and the mean number of alleles per polymorphic locus ($AP_{pop}$) was 2.52. Averaged over many plant species these values are $H_{pop} = 0.113$, a $P_{pop} = 34.2$, and an $AP_{pop} = 2.55$ (Hamrick and Godt 1989).

The relatively high level of genetic variation in $T. pratense$ may reflect the species’ wide geographic distribution and obligate outcrossing mode of reproduction. Widely distributed plant species tend to maintain more variation than narrowly distributed species; and predominantly outcrossing species tend to have more genetic diversity overall and also tend to maintain more variation within their populations than species with higher proportions of self-pollination (Hamrick and Godt 1989; Hamrick et al. 1979). The level of genetic variation in $T. pratense$ is high even when compared to other widely distributed and outcrossed species. Species with a wide geographic distribution have an $H_{pop}$ of 0.202 and $P_{pop}$ of 0.159, a $P_{s}$ of 58.9 and $P_{s}$ of 43.0, and an $AP_{s}$ of 3.2 and $AP_{s}$ of 2.0. Predominantly outcrossing species have an $H_{pop}$ of 0.167 and $H_{pop}$ of 0.124, a $P_{s}$ of 50.1 and $P_{s}$ of 35.9, and an $AP_{s}$ of 3.0 and $AP_{s}$ of 2.4. The values for $T. pratense$ are generally higher than each of these parameters. Because high levels of genetic diversity are often deliberately maintained in crop species (Hamrick and Godt 1989; Marshall 1989; Muona 1989; Weber et al.

### Table 2. Gene diversity statistics (Nei 1973, 1977) and 10 polymorphic loci in $T. pratense$

<table>
<thead>
<tr>
<th>Locus</th>
<th>$H_{exp}$</th>
<th>$G_{reg}$</th>
<th>$G_{pop}$</th>
<th>$G_{reg(pop)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPI 1</td>
<td>0.265</td>
<td>0.016</td>
<td>0.052</td>
<td>0.932</td>
</tr>
<tr>
<td>TPI 2</td>
<td>0.258</td>
<td>0.001</td>
<td>0.115</td>
<td>0.884</td>
</tr>
<tr>
<td>GDH 1</td>
<td>0.002</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>PER 1</td>
<td>0.468</td>
<td>0.001</td>
<td>0.033</td>
<td>0.966</td>
</tr>
<tr>
<td>MDH 1</td>
<td>0.419</td>
<td>0.001</td>
<td>0.006</td>
<td>0.953</td>
</tr>
<tr>
<td>GP 2</td>
<td>0.499</td>
<td>0.057</td>
<td>0.131</td>
<td>0.812</td>
</tr>
<tr>
<td>DIA 1</td>
<td>0.480</td>
<td>0.021</td>
<td>0.020</td>
<td>0.959</td>
</tr>
<tr>
<td>ME 1</td>
<td>0.473</td>
<td>0.029</td>
<td>0.038</td>
<td>0.933</td>
</tr>
<tr>
<td>LAP 1</td>
<td>0.483</td>
<td>0.000</td>
<td>0.032</td>
<td>0.968</td>
</tr>
<tr>
<td>FBP 2</td>
<td>0.361</td>
<td>0.000</td>
<td>0.063</td>
<td>0.932</td>
</tr>
</tbody>
</table>

### Table 3. Gene diversity statistics (Nei 1973, 1977) for 10 polymorphic loci in $T. pratense$

<table>
<thead>
<tr>
<th>Locus</th>
<th>$H_{exp}$</th>
<th>$G_{reg}$</th>
<th>$G_{pop}$</th>
<th>$G_{reg(pop)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPI 1</td>
<td>0.308</td>
<td>0.051**</td>
<td>0.068</td>
<td>0.881</td>
</tr>
<tr>
<td>TPI 2</td>
<td>0.268</td>
<td>0.011***</td>
<td>0.184</td>
<td>0.797</td>
</tr>
<tr>
<td>GDH 1</td>
<td>0.004</td>
<td>0.000n.s.</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>PER 1</td>
<td>0.481</td>
<td>0.015**</td>
<td>0.033</td>
<td>0.952</td>
</tr>
<tr>
<td>MDH 1</td>
<td>0.431</td>
<td>0.013**</td>
<td>0.005</td>
<td>0.982</td>
</tr>
<tr>
<td>GP 2</td>
<td>0.466</td>
<td>0.015*</td>
<td>0.230</td>
<td>0.757</td>
</tr>
<tr>
<td>DIA 1</td>
<td>0.463</td>
<td>0.023**</td>
<td>0.039</td>
<td>0.938</td>
</tr>
<tr>
<td>ME 1</td>
<td>0.537</td>
<td>0.051***</td>
<td>0.023</td>
<td>0.926</td>
</tr>
<tr>
<td>LAP 1</td>
<td>0.476</td>
<td>0.002n.s.</td>
<td>0.030</td>
<td>0.968</td>
</tr>
<tr>
<td>FBP 2</td>
<td>0.361</td>
<td>0.000n.s.</td>
<td>0.063</td>
<td>0.932</td>
</tr>
</tbody>
</table>

Total genetic variation, $H_{pop}$, was partitioned into the proportion of genetic variation found among regions, $G_{reg}$, among populations within regions, $G_{pop}$, and within populations, $G_{reg(pop)}$. 

*Chi-square test of independence. ** $P < .001; *** P < .001; n.s. = not significant.
growth and high gene flow among populations of *T. pratense*, any founder effect present when the naturalized populations were first established would be ephemeral.

Estimates of gene flow are high among populations. Based on Crow and Aoki’s (1984) gene flow equation, an $N_m$ value of 4.75 was estimated among regions. Among populations within regions $N_m = 3.36$ and between years $N_m = 2.17$. Chi-square analysis indicates that the overall $G_{st}$ value, 0.028, is significantly different from zero. Still the genetic identity between populations between years is relatively high (−0.96). These year-to-year differences are probably due to the high turnover of clover plants within naturalized populations (Vescio LV and Hamrick JL, unpublished data).

It appears that the outcrossing nature of *T. pratense*, coupled with strong flying pollinators (e.g., *Bombus* spp.) and dispersal by mammals, birds, farm machinery, and other vehicles has resulted in a high level of gene exchange via pollen between cultivated and naturalized populations of red clover. This conclusion is consistent with indirect estimates of gene flow for plant species with breeding system characteristics similar to that of red clover (Hamrick et al. 1995). Our results are also consistent with genetic marker-based studies of other crop species or their related congeners (e.g., Ellstrand and Hoffman 1985; Kirkpatrick and Wilson 1988). The potential for long-distance seed and pollen movement has maintained high levels of genetic diversity within populations and reduced genetic diversity among populations.

From the USDA/ARS/CHRU, 3420 Orchard Ave., Corvalis, OR 97330 (Hagen) and the Departments of Botany and Genetics, University of Georgia, Athens, Georgia (Hamrick). The work described in this article was supported by an NIH genetics training grant, a University of Georgia Department of Botany grant, and a Sigma Xi research grant awarded to M.J.H.

The Journal of Heredity 1998:89(2)

References


Norredo. F, 1935. Segregation data, for the *Fr3* × *Clk-h* and *Fr3* × *Clk-h* crosses. Nature 136:6077.


Received November 20, 1995

Accepted May 5, 1997

Corresponding Editor: Norman F. Weeden

**Assignment of the *Fr3* Locus to Soybean Linkage Group 9**

R. G. Palmer and X. F. Chen

A study was conducted to determine the linkage relationship between the loci *Fr3* (root fluorescence) and *Ap* (acid phosphatase) in the soybean [*Glycine max* (L.) Merr.], as well as to ascertain the linkage relationship of *Fr3* with two other loci in the soybean classical linkage group 9. Exact plant-to-plant reciprocal crosses were made between near-isogenic line *Hark-Fr3* and near-isogenic line *Clark-h* (*Kunitz trypsin inhibitor null*) and *Hark-Fr3* with cultivar Norredo. *F₂* segregation data, for the *Fr3*- *Ap* gene pair from reciprocal crosses of *Hark-Fr3* × *Clark-h*, gave a recombination value of 4.51 ± 0.55 with 1464 *F₂* plants. However, reciprocal crosses of *Hark-Fr3* × Norredo gave a recombination value of 10.25 ± 0.83 with 1344 *F₂* plants. Finally, the gene order of *Fr3*- *Ap*–*Ti–Lap1* (lecine aminopeptidase) was established for the classical linkage group 9 in soybean.

*Soybean* [*Glycine max* (L.) Merr.] linkage maps composed of morphological, protein, RFLP, RAPD, and SSR markers have been published (Akkaya et al. 1995; Palmer and Hedges 1993; Rafalski and Tingey 1993; Shoemaker and Olson 1993; Shoemaker et al. 1995; Shoemaker and Specht 1995). The cultivated soybean is considered as a diploidized tetraploid species (*2n = 2x = 40* chromosomes) (Hymowitz...