Molecular Ecology (1999) **8**, 1027–1036

Genetic structure and differentiation of *Plantago major* reveals a pair of sympatric sister species

M. MORGAN-RICHARDS*† and K. WOLFF*‡
**School of Biological Sciences, University of St Andrews, Fife KY16 9TH, UK*

Abstract

Seeds of the widespread weed *Plantago major* were collected from 10 European countries, as well as Trinidad and North America. The seed collections were from populations of two taxa which are ecologically rather than geographically separated and formally recognized as the subspecies *Plantago major* ssp. *major* and *P.m.* ssp. *intermedia* (also called *P.m.* ssp. *pleiosperma*). Eight polymorphic allozyme loci and 73 random-primed DNA fragments were scored, as well as 11 morphological characters. Complete concordance between morphological traits and genetic data provides evidence that these two taxa, although very similar, are distinct species. They are both widespread, they are broadly sympatric and capable of interbreeding. However, slight morphological and ecological differences coincide with genetic clustering of populations from widely separated locations. In addition, *P. major* and *P. intermedia* differ in their population structure: *P. intermedia* has greater genetic diversity among populations and less genetic variance within populations than *P. major*. We suggest that differences between the two species in their levels of selfing may explain the distinctive genetic structure of each species. We hypothesize a link between selfing rate and lifespan of the two taxa. *P. major* is characterized by lower genetic variation among populations, a higher rate of outcrossing, longer lifespan and production of fewer seeds per seed capsule. *P. intermedia* is more highly structured with much differentiation among populations, a higher rate of inbreeding and it often grows as an annual.

Keywords: evolution, genetic diversity, mating systems, *Plantago*, population structure, subspecies

Received 29 October 1998; revision received 4 February 1999; accepted 14 February 1999

Introduction

In the study of plant population genetics advances have recently been made in understanding the theoretical effect of inbreeding on population structure (Charlesworth *et al*. 1993, 1997). However, satisfactory explanations of empirical data are not as advanced as single models fail to explain what plants are actually doing (Bonnin *et al*. 1996; Awadalla & Ritland 1997; Liu *et al*. 1998). Genetic structure of natural populations impacts on studies of differentiation, speciation, adaptation, the evolution of mating systems and the tracing of population history.

In the past, studies have been constrained by the resolving power of the genetic techniques employed to study genetic variation. Studies using random amplification of polymorphic DNAs (RAPDs) frequently have found more genetic variation than studies of the same groups of organisms examined for allozyme variation (Hidayat *et al*. 1996; Haig *et al*. 1994; Macaranas *et al*. 1995; Burlando *et al*. 1996). In many investigations the increased ability to resolve genetic variation has revealed population structure previously unrecognized (although this is not always the case, i.e. Van Oppen *et al*. 1995).

The cosmopolitan and common weed, *Plantago major* L. has been extensively studied in the Netherlands and Denmark; thus much is known of its ecology, genetics, and reproductive biology (Mølgaard 1976; Van Dijk & Van

Correspondence: M. Morgan-Richards. Fax: +64 3-479 7584; E-mail: mary.morgan-richards@stonebow.otago.ac.nz
Present addresses: †Department of Zoology, Otago University, PO Box 56, Dunedin, New Zealand; ‡University of Newcastle, Department of Agricultural and Environmental Sciences, Newcastle upon Tyne NE1 7RU, UK.

© 1999 Blackwell Science Ltd

Delden 1981; Van Dijk 1984; 1991b; Wolff 1991a; Wolff & Schaal 1992; Wolff *et al*. 1994). *P. major* is highly inbreeding but the genus to which it belongs contains species with a wide range of mating systems, from inbreeders to obligate outcrossers. *Plantago major* is wind-pollinated and its seeds, sticky when wet, are often dispersed by animals (Kuiper & Bos 1992). The worldwide distribution of *P. major* appears to have followed human movement from Europe (the genus name means footprint, a fitting term for the prostrate growth form of *P. major*, common on paths).

Although cosmopolitan in distribution, *P. major* is locally subdivided into ecotypes. Some authors have recognized these ecotypes as separate species, other authors recognized them as subspecies or varieties, in some cases using different names. Confusion arises indirectly from the fact that the concept of a 'species' has changed over the 200 years since Linnaeus described *P. major*, and directly from the fact that the morphological differences between ecotypes are subtle. *P. intermedia* Gilibert was described in 1806 but was later relegated to a subspecies of *major* (Lange 1856) and this classification is followed by Stace (1991) in The New Flora of the British Isles. However, workers in the Netherlands referred to Pilger (1937), recognizing *P. major* ssp. *pleiosperma* (e.g. Mølgaard 1976; Van Dijk 1984; Wolff & Schaal 1992; Wolff & Morgan-Richards 1998). We are convinced that the two names (*intermedia* and *pleiosperma*) refer to the same taxon and here we use *intermedia* which has precedence.

The two *Plantago* taxa in this study are distinguished by morphological characters and habitat. *P.m.* ssp. *major* is abundant on footpaths and rough ground and often survives the winter whereas *P.m.* ssp. *intermedia* does not (where studied in Denmark, Mølgaard 1976; Kuiper & Bos 1992). Of the two subspecies, *P.m.* ssp. *major* usually has wider leaves, an acute seed capsule, usually with fewer, larger seeds per capsule (4–15). In contrast, *P.m.* ssp. *intermedia*, with narrower leaves and an obtuse seed capsule, usually produces a larger number of smaller seeds in each capsule (12–25). *P.m.* ssp. *intermedia* grows on disturbed ground, in agricultural fields, and on sites close to the sea or on river banks. In these disturbed sites *P.m.* ssp. *intermedia* usually grows as an annual (Mølgaard 1976).

Although the two subspecies of *P. major* are morphologically distinct (even if such distinctions are subtle) they are widely sympatric and capable of interbreeding and were, until this study, almost genetically indistinguishable. Allozyme studies within the Netherlands could not distinguish the two subspecies although the frequency of alleles at three loci differed significantly (Van Dijk & Van Delden 1981; Van Dijk 1984). It was suggested that the morphological characters that distinguish the two taxa may be maintained by strong selection in the face of gene flow. Chloroplast DNA (cpDNA) from seven populations of *Plantago major* provided evidence of subdivision concordant with subspecies although one population sampled was mixed (Wolff & Schaal 1992). Cellular organelles do not have the same transmission as the nuclear genome, and therefore cpDNA patterns do not always reveal the genetic patterns and population structure seen with nuclear markers (see references in Schaal *et al*. 1998). Fingerprinting of five *Plantago* species showed that, as expected for a highly selfing species, the repetitive DNA in *P. major* was not as variable as that in outcrossing species within this genus (Wolff *et al*. 1994). In addition, clear differentiation between the subspecies of *P. major* was detected. Comparison of Dutch and Scottish populations of *P. major* also revealed genetic divergence between subspecies that was stronger than genetic divergence between countries (Wolff & Morgan-Richards 1998). In this study, for the first time, material from both subspecies from throughout Europe has been genetically and morphologically examined to determine whether discrete lineages are being maintained in widespread sympatry. Examination of evolutionary relationships and current genetic structure was undertaken to help understand the partitioning of genetic variation in this taxon and resolve the taxonomic confusion.

Materials and methods

Seeds were collected from 41 wild populations of *Plantago major* from 10 countries in Europe as well as Trinidad and the USA (Table 1, and see Lyons *et al*. (1997) for more details). The plants were grown in the greenhouse under uniform conditions. When plants bore ripe seeds morphological measurements were taken. For the third leaf the width, length and petiole length was measured, and the number of teeth (coded on a scale of 1–5) and number of veins were scored. For the third inflorescence the spike and scape length were measured, the position (coded on a scale of 1–5, see Mølgaard 1976) and number of seeds per capsule were scored. For each plant the number of rosettes and inflorescences were recorded. Leaf shape was calculated by dividing leaf length by width and this measure was plotted against mean number of seeds per capsule. Statistical Package for Social Sciences (SPSS) was used for principal component analysis and discriminant analysis using the 11 morphological characters.

RAPDs and inter-simple sequence repeats (SSR)

DNA was extracted from 2 g of young leaf tissue following the method of Wolff (1996). Amplification of genomic DNA used standard polymerase chain reaction (PCR) protocols in 25 µL volumes consisting of: approximately 12 ng of genomic DNA, 1× amplification buffer (Promega); 2 mM

Table 1 Seed-collecting locations and the number of plants analysed for morphological, allozyme and random DNA characters

Location	Country	Abbreviation	Number of plants			Species: assignment based on DNA data
			morphology	allozymes	DNA	
Anavriti	Greece	Ana	7	7	5	*intermedia*
Benaki, Athens	Greece	Ben	5	5	5	*intermedia*
Ebro	Spain	Ebr	9	8	6	*intermedia*
Nybol	Denmark	Den	8	8	5	*intermedia*
Noordpolderzyl haven	Netherlands	NPZ	3	3	5	*intermedia*
Tiel	Netherlands	Tie	5	5	5	*intermedia*
Los Angeles	USA	USA	7	7	5	*intermedia*
Trinidad	Trinidad	Tri	5	2	5	*intermedia*
Austria	Austria	Aus	7	7	5	*major*
Braunschweig	Germany	Ger	6	6	5	*major*
Belgium	Belgium	Bel	4	4	5	*major*
Belp	Switzerland	Swt	4	4	4	*major*
Olympus	Greece	Oly	5	5	5	*major*
Segovia	Spain	Seg	6	6	5	*major*
Aston Hill	Britian	Ast	11	11	10	*intermedia & major*
Bottesford	Britian	Bot	2	3	4	*major*
Eskdalemuir	Britian	Esk	6	7	4	*major*
East Malling	Britian	Mal	7	7	3	*major*
Harwell	Britian	Har	5	6	4	*intermedia & major*
Bakewell	Britian	Bak	17	22	22	*major*
Leatherhead	Britian	Lea	3	3	3	*major*
Lullington	Britian	Lul	19	23	14	*major*
Loch Maddy	Britian	LMa	5	5	1	*major*
Saint Osyth	Britian	Osy	9	11	8	*major*
Scaftworth	Britian	Sca	7	9	11	*major*
Sibton	Britian	Sib	7	9	5	*major*
Strath Vaich	Britian	StV	4	4	4	*major*
Totley	Britian	Tot	8	10	13	*major*
Wray	Britian	Wra	5	5	5	*major*

MgCl; 100 μM of each dNTP; 0.5 U *Taq* (Promega); and 0.2 pmol of primer. The random primers used were 12 RAPD primers (Operon B4, B10, B11, B13, G2, G3, G5, G10, G12, G14, G18, G19) and five anchored inter-SSR primers (UBC 811, 827, 830, 847, 889) (Zietkiewicz *et al*. 1994). The thermal cycling for the anchored inter-SSR primers was as follows: 45 cycles at 94 °C for 30 s, 52 °C for 45 s, 72 °C for 2 min. The protocol for the RAPD primers was as follows: 45 cycles of 94 °C for 15 s, 36 °C for 45 s (with a ramp rate of 0.4 °C/s), 72 °C for 1.5 min. The amplified products were separated by electrophoresis in a 1.4% agrose gel, stained with ethidium bromide and photographed under UV illumination. Only DNA fragments of high concentration with a length between 0.25 kb and 1.5 kb were used as markers (Pérez *et al*. 1998). Scoring for the presence or absence of DNA fragments was aided by the use of a 1 kb DNA ladder and a control sample. Every set of amplification reactions (96-well PCR tray) included five reactions using DNA from a 'control' plant (*P.m.* ssp. *major* ISP 1994 329). Before any DNA fragment was used as a marker in the analysis we ensured that its amplification or nonamplification (presence or absence) was consistent over the five PCR products of the control sample in each reaction tray. A lack of reproducibility of fragments produced by the RAPD method has recently been highlighted but with careful use of controls and scoring blind we agree with Pérez *et al*. (1998) that it is a suitable method for the identification of species and strains.

Allozymes

We surveyed eight polymorphic allozyme loci (number of alleles in brackets): *Got*-1(3), *Got*-2(3), *Gpi*-1(2), *6Pgd*-1(2), *6Pgd*-2(2), *6Pgd*-3(2), *Pgm*-1(3), *Shdh*(4). The tissue preparation, buffers and staining methods followed Van Dijk & Van Delden (1981) and Shaw & Prasad (1970) for starch gel electrophoresis. *Est*-4 was not scored although it was previously shown to be polymorphic. We did not observe three alleles recorded in previous studies of *P. major*: slow

Me-1, fast *Acph*-2 and the null allele of *6Pgd*-2. However, we did observe three alleles not seen before: slow *6Pgd*-1, fast *6Pgd*-2, and slow *6Pgd*-3.

Population genetics

Cluster analysis using neighbour-joining (Saitou & Nei 1987) with PHYLIP (Felsenstein 1993) and printed using TreeView was performed with: (i) Euclidean distance (calculated with Arlequin (Schneider *et al*. 1997)) from the RAPD data; and (ii) arc distance (Cavalli-Sforza & Edwards 1967) calculated from the allozyme data using BIOSYS (Swofford & Selander 1981). The choice of distance measure had little effect on the structure of the resulting tree.

F-statistics were estimated using the program FSTAT (Weir 1990) with the allozyme data set, including bootstrap by loci to generate 95% confidence intervals. F_{IS} values for each population were compared using a Mann–Whitney *U*-test to determine if *P.m.* ssp. *intermedia* was significantly less outcrossing than *P.m.* ssp. *major* (one-tailed test). Measures of population heterozygosity (H_E) were calculated from the allozyme data using GelJock (Porter 1995), and genetic diversity (D_{ST}) (Nei 1973) calculated as $(H_S.F_{ST})/1 - F_{ST}$. Total variability was expressed as $H_T = H_S + D_{ST}$. The RAPD and inter-SSR data were analysed with AMOVA (Excoffier *et al*. 1992) and Arlequin (Schneider *et al*. 1997). Randomised permutations we used to compare Φ_{ST} estimates for the two subspecies. Assignment of each population to one or other subspecies was randomised, keeping the number of populations constant and Φ_{ST} calculated for each taxon. Although RAPD data are dominant we used programs designed for haplotype frequencies because *P. major* is highly inbreeding; thus we assumed that the low level of hidden heterozygosity would have a negligible effect on the calculations. Stewart & Excoffier (1996) showed that Φ_{ST} estimates were little altered when 'allele' frequencies were estimated assuming a high selfing rate (0.9) and incorporated into AMOVA calculations.

Results

Morphological and genetic data were scored for each plant without knowledge of the morphological or ecological characteristics of the populations from which the seeds originated. During data analysis plants were assigned to one of two taxa using both genetic and morphological characters.

Morphology

Principal component analysis was used without prior classification of the plants, and using all 11 morphological characters this method did not produce clear and distinct clusters of plants. However, when the plants were coded into two groups based on the genetic results (see neighbour-joining tree below), then simple scatter plots of some morphological characters produced distributions that were almost nonoverlapping (Fig. 1). Discriminative analysis using the morphological characters assigned the plants to the two genetic groups with 100% accuracy. The mean number of seeds per capsule is the character that most successfully separates the plants into two groups, and used in combination with spike length or leaf width it can usually distinguish the two subspecies in our study.

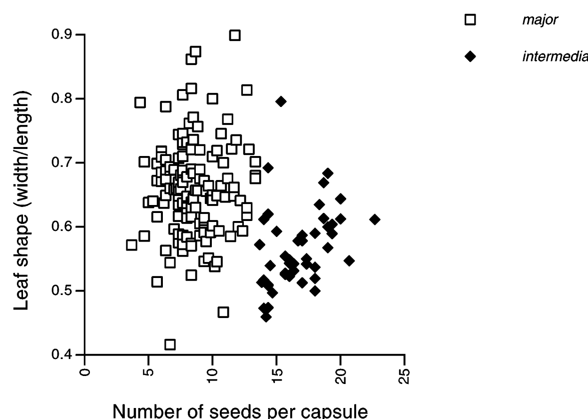

Fig. 1 The two taxa *Plantago major* ssp. *major* and *Plantago major* ssp. *intermedia* can be distinguished based on the relationship between number of seeds per capsule and leaf shape (width/length).

The morphological and genetic data suggested that from the majority of locations a single subspecies had been collected. However, from the morphological analysis it appeared that two of the 112 seeds collected from Britain were within the *P.m.* ssp. *intermedia* group: one plant from Harwell and one from Aston Hill; the other plants from these populations are of the subspecies *major*. Two plants from Sibton with fewer than 12 seeds per capsule as expected for *P.m.* spp. *major* had uncharacteristically narrow leaves and a small scape:spike ratio.

RAPD and inter-SSR DNA fragments

Between one and seven DNA fragments were scored as present or absent for each random primer. Only polymorphic markers were recorded; these are 22 anchored inter-SSR markers and 51 RAPD markers (data available from authors upon request). Twenty-five of these 73 markers (34.2%) were present in less than 10% or in more than 90% of the samples. Thirteen (17.8%) of the markers were specific (or nearly so) to one of the two subspecies (Table 2). The majority of these diagnostic fragments were amplified only in *P.m.* ssp. *major.*

Table 2 Thirteen markers that are almost diagnostic for the two *Plantago major* subspecies in Europe showing the number of plants with the marker out of the total number of plants scored

Marker	*major*	*intermedia*
1	138/138	1/43
2	93/138	0/43
3	132/136	0/41
4	133/137	1/42
5	136/137	1/43
6	116/136	2/43
7	131/138	7/43
8	21/138	42/43
9	0/138	35/43
10	133/137	10/42
11	132/138	1/42
12	138/138	2/43
13	133/138	5/43

Pairwise Euclidean distances were calculated from the 73 polymorphic markers and used to cluster plants that were genetically most similar. Two groups of plants were clearly distinguished by the neighbour-joining analysis of all RAPD and inter-SSR markers (Fig. 2). One group of 43 plants was identified as *P.m.* ssp. *intermedia.* Within this clade, plants cluster first with plants that were collected from the same location. The exception was one plant from LA airport (USA) which was more similar to plants from Trinidad than to plants from the same collection site. The plants from the nine other locations formed nine monophyletic clades within *P.m.* ssp. *intermedia.* Within the second major clade (lower clade Fig. 2) all the *P.m.* ssp. *major* plants cluster but with much less structure than seen within *P.m.* ssp. *intermedia.* Although for some collection locations of *P.m.* ssp. *major* the plants clustered together (e.g. Belgium), for other locations the plants were as genetically similar to plants from elsewhere as they were to geographically adjacent plants.

All but two of the plants from Britain cluster with the *P.m.* ssp. *major* samples; the exceptions are from Harwell and Aston Hill. These are the same plants whose morphological characters placed them with the *P.m.* ssp. *intermedia* plants. Thus, we have clear concordance of morphology and genetics. These two British *P.m.* ssp. *intermedia* have unique genotypes which is not what would be expected if these plants represented seeds that had been mislabelled. Instead of clustering with other populations sampled, they appear to represent distinct *P.m.* ssp. *intermedia* populations, sympatric with *P.m.* ssp. *major* in Britain. The rarity of British *P.m.* ssp. *intermedia* in this study is probably not an accurate reflection of their occurrence in Britain but due to collecting bias which concentrated on sites favoured by *P.m.* ssp. *major*.

The *P.m.* ssp. *major* plants genetically most similar to *P.m.* ssp. *intermedia* were collected in Sibton. These plants show character states characteristic of *P.m.* ssp. *intermedia* for nine DNA markers, three are shared-absence of fragments and six are shared-presence of fragments. Why gene flow between these two subspecies should be detected at a single site and why at Sibton, are questions that can be addressed in further studies.

Population structure within *P. major* as a whole was estimated using AMOVA. Variation between subspecies (Φ_{ST}) describes 46% of the total genetic variance within the combined data (Table 3). In order to compare estimates of population structure within each subspecies, the RAPD and inter-SSR data were divided using the two clades in Fig. 2. To aid comparisons most of the *P.m.* ssp. *major* populations from Britain were excluded to make the two groups more similar in both their geographical coverage and sample size. *P.m.* ssp. *intermedia* was represented by 10 populations (Denmark, Spain, USA, Trinidad, two sites from the Netherlands, UK and Greece), and *P.m.* ssp. *major* was represented by nine populations [Austria, Greece, Spain, Germany, Belgium, Switzerland, and three from the UK (Bakewell, Sibton, Wray)]. Using AMOVA (Excoffier *et al*. 1992) greater genetic variation among populations was detected within *P.m.* ssp. *intermedia* ($\Phi_{ST} = 0.672$) than within *P.m.* ssp. *major* ($\Phi_{ST} = 0.397$). Random permutations of the populations reveals that the Φ_{ST} for *P.m.* ssp. *major* is significantly lower than expected if the assignment of populations to the two taxa was random. In *P.m.* ssp. *intermedia* only a third of the genetic variance is within populations while for *P.m.* ssp. *major* almost two-thirds of the genetic variance is within population (Table 3). As genetic variation was partitioned between populations to a greater extent in *P.m.* ssp. *intermedia* it can be viewed as the more highly structured of the two taxa.

Allozymes

Of the eight polymorphic loci studied none had alleles that differentiated the two subspecies. There were alleles unique to Greek *P.m.* ssp. *intermedia* (*Gpi* fast), Spanish *P.m.* ssp. *intermedia* (*6Pgd*-1 fast), and rare alleles found only in *P.m.* ssp. *major* (*Got*-2 slow, *Skdh* med-fast, *6Pgd*-2 fast, *6Pgd*-3 slow, *Pgm* fast and slow). Genotype and allele frequencies are not presented here but are available on request. A neighbour-joining tree from pairwise genetic distance data revealed little structure (not shown), the lack of resolution being due to low levels of variation and the widespread sharing of alleles, as seen in previous allozyme studies of this species (Van Dijk & Van Delden 1981). The frequency of many alleles differed between the two subspecies but due to the high level of inbreeding, unequal sample sizes and unequal ratio of the two taxa in our data set (almost four-times more *P.m.* ssp. *major* than

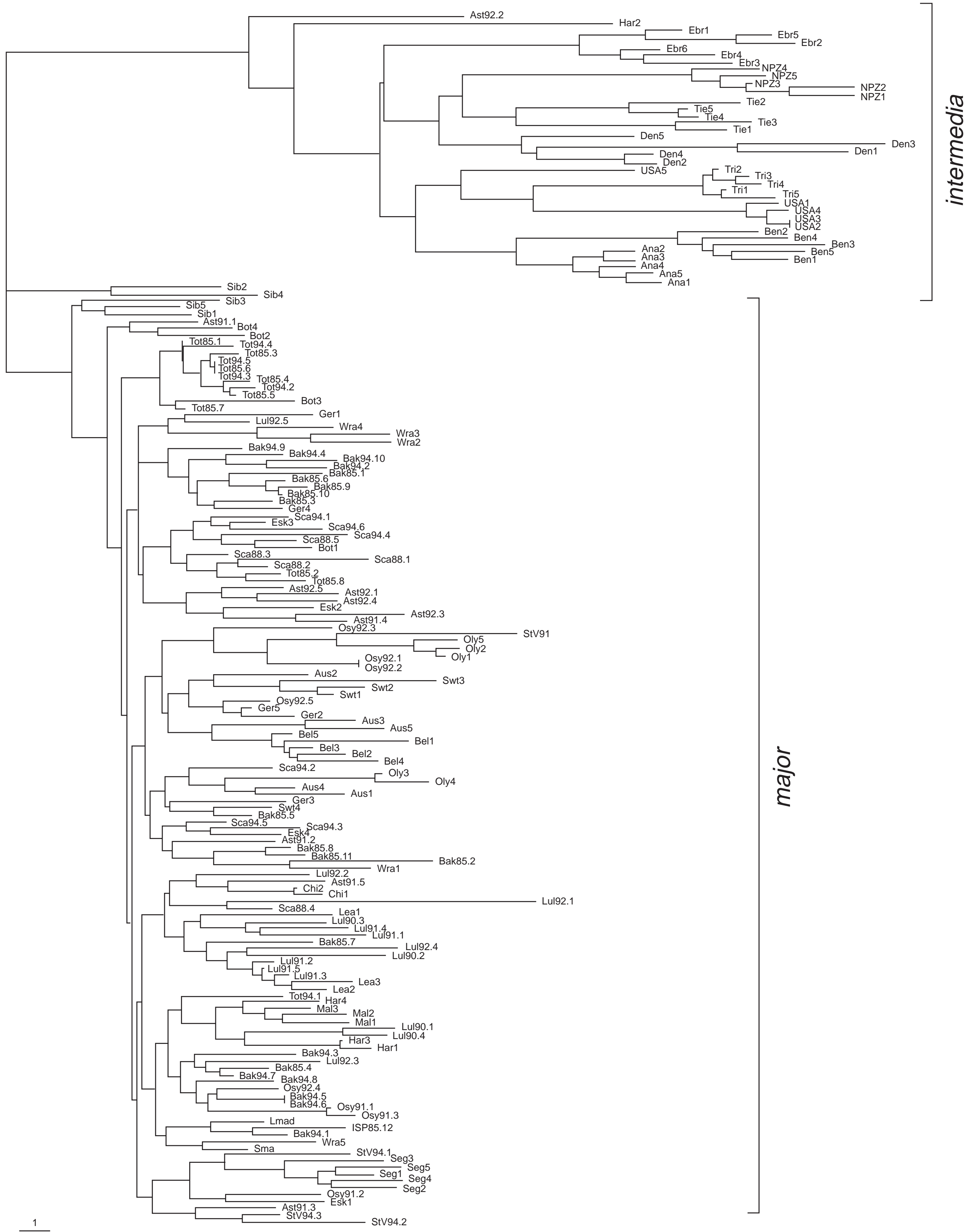

Fig. 2 Neighbour-joining tree (Saitou & Nei 1987) showing the genetic relationship between *Plantago major* populations using pairwise Euclidean distances calculated from 73 RAPD and inter-SSR markers (Schneider *et al*. 1997).

Table 3 Genetic variation within and among *Plantago major* subspecies estimated using 73 random genetic markers and AMOVA (Excoffier *et al*. 1992), using a nested analysis of the variance components for both subspecies. Φ_{ST} is the correlation of markers within populations relative to the whole data set, Φ_{SC} is the correlation of markers within populations relative to the subspecies to which it belongs, Φ_{CT} is the correlation of markers within subspecies relative to the whole data set

Variance component		Φ	
Among subspecies	46.30%	Φ_{ST}	0.763
Among pops within subspecies	30.00%	Φ_{SC}	0.559
Within populations	23.70%	Φ_{CT}	0.463
P.m. ssp. *major* only		Φ_{ST}	0.397
P.m. ssp. *intermedia* only		Φ_{ST}	0.672

P.m. ssp. *intermedia*), these frequency differences cannot be viewed as significant.

The allozyme data for the *Plantago* samples were divided into two groups, corresponding to the two subspecies, and by excluding some of the *P.m.* ssp. *major* populations from Britain these groups were made similar in both their geographical coverage and sample size. The nine populations of *P.m.* ssp. *intermedia* were from Britain, Denmark, Spain, USA, Trinidad and two sites in Greece and the Netherlands. The nine populations of *P.m.* ssp. *major* were from Austria, Belgium, Greece, Germany, Switzerland, Spain, and three sites in Britain. Thus *F*-statistics can be directly compared for the similarly geographically widespread groups (Table 4). Significantly more genetic variation was described by among-population differences in *P.m.* ssp. *intermedia* ($F_{ST} = 0.777$) than by among-population differences in *P.m.* ssp. *major* ($F_{ST} = 0.232$). In contrast, *P.m.* ssp. *intermedia* had a lower mean expected heterozygosity (H_S) than *P.m.* ssp. *major*. The distribution of within-population allozyme variability (H_E) for these two taxa revealed that *P.m.* ssp. *major* had a greater range of mean expected heterozygosity per population (Fig. 3). Total genetic diversity (H_T) of *P.m.* ssp. *intermedia* was slightly higher than total genetic diversity in *P.m.* ssp. *major* (Table 4). The level of inbreeding within each population (F_{IS}) revealed *P.m.* ssp. *intermedia* to have a significantly greater rate of selfing than *P.m.* ssp. *major* ($P < 0.10$). The F_{ST} value for *P.m.* ssp. *intermedia* was skewed down by a single locus (*6Pgd*-1) in a single population (Ebro, Spain); removal of this population reduces our estimation of outcrossing in *P.m.* ssp. *intermedia* to zero, significantly lower than the 10.3% rate of outcrossing in *P.m.* spp. *major*.

Discussion

In contrast to some earlier genetic studies of *Plantago major* (Van Dijk & Van Delden 1981), this study has resolved enough genetic variation to unequivocally distinguish the two subspecies and to resolve population structure. Our analysis clearly separates the plants into two taxa, showing complete concordance of the morphological and genetic data sets. In addition, within *P.m.* ssp. *intermedia* there is concordance of genetic similarity and physical proximity. Both the allozyme and DNA data provide estimates of F_{ST} that indicate that these two taxa have quite distinct population structures with much more diversity among populations of *P.m.* ssp. *intermedia* than seen among populations of *P.m.* ssp. *major*.

Species or ecotypes?

There is no consistency in the way botanists use subspecies and varieties to classify organisms. Frequently, subspecies are used to define regionally separated races that are morphologically differentiated (McDade 1995) rather than to classify ecotypes which are separated by habitat (Stace 1989; Hamilton & Reichard 1992; Stuessy 1994), as is the case here. We consider that *P. major* and *P. intermedia* should be regarded as separate species in agreement with Gilibert 1809 and shall refer to them thus for the remainder of this study. Our data suggest that the two taxa are genetically differentiated even in sympatry, and have distinct patterns of genetic population structure. There is no absolute measure of genetic difference that determines what are separate species. However, comparison with similar studies and examination of genetic evidence in tandem with other information such as the ecology, morphology and mating system allows for an educated view of the independence of the taxa, their cohesion and their potential for separate evolutionary fates. This study can, for instance, be compared with a study of Howard *et al*. (1997) in which they compared two oak species using RAPDs. The two oaks are considered to be good species that can be discriminated by morphological characters and are placed in two separate subsections of the white oaks. Howard *et al*. (1997) used

Table 4 Genetic variation of eight polymorphic allozyme loci within and among populations of *Plantago major* using Weir & Cockerham's (1984) method

	F_{IS} (SD)	F_{IT}	F_{ST} (SD)	D_{ST}	H_S	H_T
P. m. ssp. *intermedia*	0.929 (0.031)	0.675	0.777 (0.024)	0.139	0.04	0.179
P. m. ssp. *major*	0.818 (0.021)	0.768	0.232 (0.037)	0.024	0.08	0.104

© 1999 Blackwell Science Ltd, *Molecular Ecology*, 8, 1027–1036

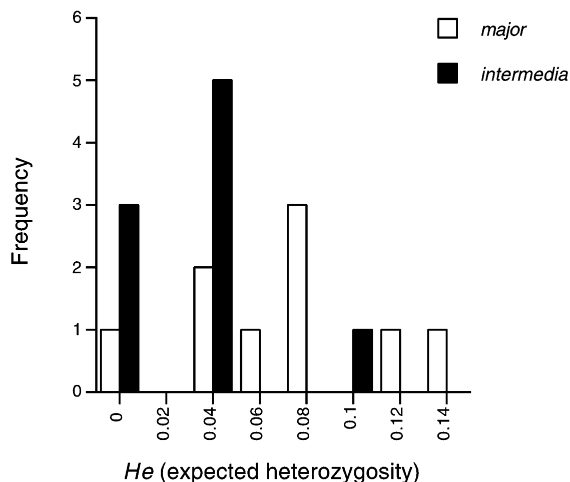

Fig. 3 The distribution of genetic diversity within populations of the two taxa *Plantago major* and *P. intermedia* based on allozyme data from eight polymorphic loci.

700 different primers and only found 14 markers with six primers that were species specific. The two *Plantago* taxa studied here have as many markers that discriminate them in widespread sympatry, in addition to the morphological and ecological characters.

Having said that we consider these two species to be genetically distinct, we recognize that clustering of these plants into two groups could be a genetic artefact rather than evidence of common ancestry. It is possible that the markers that distinguish the two species in this study are all linked to selected traits (such as width of leaves and the number of seeds per capsule). The maintenance of ecotypes via selection against hybrids was suggested by Van Dijk & Van Delden (1981) who used allozymes to study Dutch populations of these species. Selection against ecotype hybrids has been documented in other plant species (e.g. Nagy 1997). With 73 random markers analysed, most of which probably originate in the nuclear DNA (Lorenz *et al*. 1994), it seems unlikely that the structure observed could be dominated by markers linked to traits under strong selection because recombination would break up linkage groups. However, because these species have high selfing rates, it may only require 13 or 14 markers that are loosely linked to alleles under strong selection to produce quite different frequencies of these markers, and create the divergence observed (Charlesworth *et al*. 1997). If 13 markers in our study were linked to the number of seeds per capsule then one might see genetic clustering of plants that concurred with taxonomy. Thus, background selection and selective sweeps could reduce genetic variability through hitch hiking or linkage of markers to alleles under negative or positive selection and thus produce random clustering of populations. However, we found specific markers in each species, not on a local scale but on a European scale, which could well be specific on a world scale. Linkage of all diagnostic DNA markers would have to be maintained in populations from Greece to Trinidad (over many generations) for the clustering of the DNA phenotypes to concur with taxonomic groups. The selection for ecotypes as an explanation for our two genetic clusters is being tested with crossing experiments to follow the transmission of morphological traits and RAPD and inter-SSR markers in hybrid plants.

Compelling evidence that there are two taxa represented in our study does not only rely on 13 diagnostic markers, but also on the dramatic difference in the genetic structure of the two groups (Fig. 2). The difference in the Φ_{ST} estimates for the two species is greater than expected from a random division of this data into two taxa. We therefore feel that the consistency over a wide geographical range indicates that each taxon is distinct with the genetic cohesion of a full species. Hybrids that are observed in the field and produced in the greenhouse (Mølgaard 1976), and gene flow seen at Sibton, confirm that occasional hybridization does occur. Potentially lower fitness of hybrids might restrict gene flow between the two species and maintain their independence. Van Dijk & Van Delden (1981) noted deviations from Mendelian ratios in F_1 hybrids and male sterility of some F_2 hybrids. The fitness of hybrids will be surveyed in future studies.

Genetic diversity and population structure

Many studies endeavour to determine the relative role of forces that shape the level of genetic variation in natural populations. However, it may not be possible to distinguish between patterns of genetic structure that result from current mating systems and those patterns that have arisen from the population and evolutionary history of the organism (Schaal *et al*. 1998). When less variation is found than expected, distinguishing between population bottlenecks and background selection is not always possible (Awadalla & Ritland 1997; Liu *et al*. 1998). Or, if more variation is found within a subset of populations of inbreeding plants, then a posthoc historical scenario is suggested to explain the results (Bonnin *et al*. 1996).

P. intermedia has greater genetic diversity among populations than is observed among populations of *P. major* (measured using both the DNA and allozyme data). Although the level of genetic diversity within populations of *P. major* is on average greater than within *P. intermedia* populations, the total gene diversity of the species is greater in *P. intermedia*. Four explanations for the difference in genetic diversity and population structure observed in the two species can be suggested. These explanations are not mutually exclusive. In fact, to explain both the

Table 5 Estimates of outcrossing rates calculated using the fixation index (t_F) and multilocus estimation from a progeny analysis (t_m) for two Dutch populations of *Plantago intermedia* and *P. major* (1 & 2) (Wolff 1991b) and results from this study (3)

	F_{IS}	t_F	t_m
P. intermedia (1)	0.889	0.059	0.013
P. intermedia (2)	0.948	0.027	0
P. intermedia (3)	0.927	0.037	
P. major (1)	0.747	0.145	0.076
P. major (2)	0.750	0.143	0.016
P. major (3)	0.818	0.103	

difference in degree of variation and the different structure of the variation may well require more than one causal factor.

1. *P. major* may have been derived from *P. intermedia* and thus represent a subset of the genetic variation within *P. intermedia*.
2. *P. major* may have lost variation due to a more recent or more severe bottleneck and/or due to background selection and selective sweeps.
3. *P. major* may have a greater rate of outcrossing than *P. intermedia*.
4. The rate of molecular evolution in *P. major* might be slower than within *P. intermedia*.

The distinct population genetic structure seen here could be explained by just a small difference in rate of outcrossing because the frequency of outcrossing is the most important determinant of population genetic structure, affecting both genetic diversity within populations and genetic diversity among them (Barrett & Harder 1996). Previous studies of both natural populations and offspring experiments of *P. major* and *P. pleiosperma* (= *P. intermedia*) have found significant variation in their levels of selfing (Table 5), in agreement with this study. As seen in other inbred plant species (Wolff & Schaal 1992; Charlesworth *et al*. 1993; Hamrick & Godt 1996; Liu *et al*. 1998), a higher rate of outcrossing in *P. major* would result in homogenizing of the genetic variation compared to *P. intermedia* where a higher rate of inbreeding would lead to more marked differentiation among populations. Local selection would tend to increase diversity between populations more in the inbreeding taxon than the outbreeder (Charlesworth *et al*. 1997).

The change in level of outcrossing in *Plantago,* in tandem with slight ecological shifts suggests that these changes, and perhaps others, as yet unrecorded, have changed the selection pressure in favour of greater outcrossing in *P. major* (or alternatively in favour of greater inbreeding in *P. intermedia*). Such changes provide the opportunity to identify factors that promote and maintain sex (Hurst & Peak 1996). In this example lifespan may be a key component in determining the rate of outcrossing. *P. intermedia* is able to complete its life cycle in one year (summer) and is less hardy in the winter than *P. major* which often occurs as a perennial (Mølgaard 1976; Kuiper & Bos 1992). We believe that viewing *P. major* and *P. intermedia* as independent taxa is essential for the development of ideas towards understanding what we have observed.

Acknowledgements

We are pleased to thank Alan Davison and his collaborators who provided many of the *Plantago* seeds, Clive Nuttman who counted most of the seeds, and Jaime Estrella, Andy Lowe, Juliet Perry-James, Jenny Gleason, and David Forbes for technical and social support and Richard Abbort for sharing his laboratory. This paper was greatly improved by suggestions from Deborah Charlesworth, three anonymous reviewers and Loren Rieseberg. The work was finacially supported by a NERC grant Gr9/2396, and the writing of this paper was supported by the Marsden Fund (PVT601 to MMR).

References

Awadalla P, Ritland K (1997) Microsatellite variation and evolution in the *Mimulus guttatus* species complex with contrasting mating systems. *Molecular Biology and Evolution*, **14**, 1023–1034.

Barrett SCH, Harder LD (1996) Ecology and evolution of plant mating. *Trends in Ecology and Evolution*, **11**, 73–79.

Bonnin I, Huguet T, Gherardi M, Prosperi J, Olivieri I (1996) High level of polymorphism and spatial structure in a selfing plant species, *Medicago truncatula* (Leguminosae), shown using RAPD markers. *American Journal of Botany*, **83**, 843–855.

Burlando B, Arillo A, Spano S, Machetti M (1996) A study of the genetic variability in populations of the European woodcock (*Scolopax rusticola*) by random amplification of polymorphic DNA. *Italian Journal of Zoology*, **63**, 31–36.

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: Models and estimation procedures. *Evolution*, **21**, 550–570.

Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics*, **134**, 1289–1303.

Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research*, **70**, 155–174.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distance among DNA haplotypes: applications to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Felsenstein J (1993) PHYLIP 3.5 (available from joe@genetics.washington.edu). University of Washington.

Haig SM, Rhymer JM, Heckel DG (1994) Population differentiation in randomly amplified polymorphic DNA of red-cockaded woodpeckers *Picoides borealis*. *Molecular Ecology*, **3**, 581–593.

Hamilton CW, Reichard SH (1992) Current practice in the use of subspecies, variety, and forma in the classification of wild plants. *Taxon*, **41**, 485–498.

Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. *Philisophical Transactions of the Royal Society of London B*, **351**, 1291–1298.

Hidayat P, Phillips TW, Frenchconstant RH (1996) Molecular and morphological characters discriminate *Sitophilus oryzae* and *Sitophilus zeamais* (Coleoptera, Curculionidae) and confirm reproductive isolation. *Annuals of the Entomological Society of America*, **89**, 645–652.

Howard DJ, Preszler RW, Williams J, Fenchell S, Boecklen WJ (1997) How discrete are oak species? Insights from a hybrid zone between *Quercus grisea and Quercus gambelii*. *Evolution*, **51**, 747–755.

Hurst LD, Peak JR (1996) Recent advances in understanding of the evolution and maintenance of sex. *Trends in Ecology and Evolution*, **11**, 46–52.

Kuiper PJC, Bos M (1992) *Plantago: a Multidisciplinary Study*. Springer, Berlin.

Lange H (1856) *Handbook Danske Flora*. 2nd edn. Reitzel, Copenhagen, Denmark.

Liu F, Zhang L, Charlesworth D (1998) Genetic diversity in Leavenworthia populations with different inbreeding levels. *Proceedings of the Royal Society of London B*, **265**, 293–301.

Lorenz M, Weine A, Borner T (1994) DNA fragments of organellar origin in random amplified polymorphic DNA (RAPD) patterns of sugar beet (*Beta vulgaris* L.). *Theoretical and Applied Genetics*, **88**, 775–779.

Lyons TM, Barnes JD, Davison AW (1997) Relationships between ozone resistance and climate in European populations of *Plantago major* L. *New Phytologist*, **136**, 503–510.

Macaranas JP, Mather P, Hoeben P, Capra M (1995) Assessment of genetic variation in wild populations of the Redclaw Crayfish (*Cherax quadricarinatus*, Vonmartens 1868) by means of allozyme and RAPD–PCR Markers. *Marine and Freshwater Research*, **46**, 1217–1228.

McDade LA (1995) Species concepts and problems in practice. Insight from botanical monographs. *Systematic Botany*, **20**, 606–622.

Mølgaard P (1976) *Plantago major* ssp. *major* and ssp. *pleiosperma*. Morphology, biology and ecology in Denmark. *Botany Tidsskrift Bd*, **71**, 31–56.

Nagy ES (1997) Selection for native characters in hybrids between two locally adapted plant species. *Evolution*, **51**, 1469–1480.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Acadamy of Sciences of the USA*, **70**, 3321–3323.

Pérez T, Albornoz J, Domínguez A (1998) An evaluation of RAPD fragment reproducibility and nature. *Molecular Ecology*, **7**, 1347–1358.

Pilger R (1937) *Das Pflanzenreich Regni vegetabilis conspectus*. Part bvev. 269 Plantaginaceae. Neudruck, p. 466.

Porter A (1995) GelJock. Department of Biology, Bowling Green St. University, Bowling Green, OH 43403–0212, USA.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406–425.

Schaal BA, Hayworth DA, Olsen KMJT, Rauscher JT, Smith WA (1998) Phylogeographic studies in plants: problems and prospects. *Molecular Ecology*, **7**, 465–474.

Schneider S, Kueffer J-M, Roessli D, Excoffier L (1997) Arlequin, Version 1.1: Software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Shaw CR, Prasad R (1970) Starch gel electrophoresis of enzymes—A compilation of recipes. *Biological Genetics*, **4**, 297–320.

Stace CA (1989) *Plant Taxomony and Biosystematics*. Edwards Arnold, London.

Stace CA (1991) *New Flora of the British Isles*. Cambridge University Press, New York.

Stewart CN, Excoffier L (1996) Assessing population genetic structure and variability with RAPD data: application to *Vaccinium macrocarpon* (American cranberry). *Journal of Evolutionary Biology*, **9**, 153–171.

Stuessy TF (1994) *Case Studies in Plant Taxonomy*. Columbia University Press, New York.

Swofford DL, Selander RB (1981) BIOSYS-1; A FORTRAN program for the comprehensive analysis of electrophoretic data in population genetics and systematics. *Journal of Heredity*, **72**, 281–283.

Van Dijk H (1984) Genetic variability in *Plantago* species in relation to their ecology. Part 2. Quantitative characters and allozyme loci in *P. major*. *Theoretical and Applied Genetics*, **68**, 43–52.

Van Dijk H, Van Delden W (1981) Genetic variability in *Plantago* species in relation to their ecology. Part 1. Genetic analysis of the allozyme variation in *P. major* subspecies. *Theoretical and Applied Genetics*, **60**, 285–290.

Van Oppen M, Olsen J, Stam W (1995) Genetic variation within and among North-Atlantic and Baltic populations of the benthic alga *Phycodrys rubens* (Rhodophyta). *European Journal of Phycology*, **30**, 251–260.

Weir BS (1990) Intraspecific differentiation. In: *Molecular Systematics* (eds Hillis DM, Moritz C), pp. 385–405. Sinauer Associates, USA.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Wolff K (1991a) Genetic analysis of morphological variability in three *Plantago* species with different mating systems. *Theoretical and Applied Genetics*, **81**, 111–118.

Wolff K (1991b) Analysis of allozyme variability in three *Plantago* species and a comparison to morphological variability. *Theoretical and Applied Genetics*, **81**, 119–126.

Wolff K (1996) RAPD analysis of sporting and chimerism in chrysanthemum. *Euphytica*, **89**, 159–164.

Wolff K, Morgan-Richards M (1998) PCR markers distinguish *Plantago major* subspecies. *Theoretical and Applied Genetics*, **96**, 282–286.

Wolff K, Rogstad SH, Schaal B (1994) Population and species variation of minisatellite DNA in *Plantago*. *Theoretical and Applied Genetics*, **87**, 733–740.

Wolff K, Schaal B (1992) Chloroplast DNA variation within and among five *Plantago* species. *Journal of Evolutionary Biology*, **5**, 325–344.

Zietkiewicz E, Rafalski A, Labuda D (1994) Genome fingerprinting by simple sequence repeats (SSR)-anchored PCR amplification. *Genomics*, **20**, 176–183.

This study is part of a research programme coordinated by Kirsten Wolff to investigate the effect of mating systems on genetic structure and diversity using the genus *Plantago* as a model system. The laboratory work was undertaken at the University of St Andrews by Mary Morgan-Richards, a postdoctoral fellow and population geneticist. Mary is now at the University of Otago, New Zealand, with Marsden funding, to compare six hybrid zones within a single weta species. Kirsten Wolff is continuing her studies of *Plantago* species and the evolution of repetitive DNA in plants at the University of Newcastle.