Molecular Ecology (2007) **16**, 2317–2332 doi: 10.1111/j.1365-294X.2007.03314.x

Population structure and genetic diversity in tristylous *Narcissus triandrus*: insights from microsatellite and chloroplast DNA variation

KATHRYN A. HODGINS and SPENCER C. H. BARRETT
Department of Ecology & Evolutionary Biology, University of Toronto, 25 Willcocks Street, Toronto, ON M5S 3B2, Canada

Abstract

We investigated cpDNA sequence and nuclear microsatellite variation among populations of the wild daffodil *Narcissus triandrus* to examine the role of historical vs. contemporary forces in shaping population structure, morphological differentiation and sexual-system evolution. This wide-ranging heterostylous species of the Iberian Peninsula is largely composed of two allopatric varieties (vars. *cernuus* and *triandrus*), and populations with either stylar trimorphism or dimorphism. Dimorphic populations only occur in var. *triandrus*, are mainly restricted to the northwestern portion of the species range, and uniformly lack the mid-styled morph (M-morph). Chloroplast DNA (cpDNA) sequence variation revealed strong geographical structuring and evidence for a fragmentation event associated with differentiation of the two varieties. In var. *triandrus*, population fragmentation, restricted gene flow and isolation-by-distance were also inferred. Significant differences in genetic diversity and population structure between the two varieties likely reflect historical and contemporary differences in demography and gene flow among populations. Discordance between cpDNA markers and both microsatellite and morphological variation indicate that hybridization has occurred between the two varieties at contact zones. There were no differences in genetic diversity or population structure between dimorphic and trimorphic populations, and chloroplast haplotypes were not associated with either sexual system, indicating transitions in morph structure within each maternal lineage. M-morph frequencies were positively correlated with differentiation at microsatellite loci, indicating that the evolutionary processes influencing these neutral markers also influence alleles controlling the style morphs.

Keywords: admixture, heterostyly, *Narcissus*, phylogeography, population structure

Received 4 September 2006; revision received 13 December 2006; accepted 5 February 2007

Introduction

The distribution of neutral genetic variation at marker loci enables inferences to be made concerning phylogeography, historical patterns of dispersal and migration, and demographical history. However, historical events not only influence neutral genetic markers but they can also affect the distribution of alleles governing traits that impact fitness. Indeed, a major challenge for understanding the maintenance of adaptive genetic variation is to determine the relative importance of historical contingency and contemporary selective regimes in structuring variation patterns. The influence of historical and demographic forces on neutral variation has recently been investigated in order to understand geographical variation of ecologically important traits in various species (e.g. Gubitz *et al*. 2000; Garcia-Paris *et al*. 2003). However, investigations of geographical variation in reproductive systems have largely been conducted from an ahistorical perspective, with a focus on the microevolutionary forces maintaining variation in contemporary populations (reviewed in Barrett *et al*. 2001).

Plant sexual polymorphisms (e.g. heterostyly and dioecy) have provided insights into the ecology and evolution of populations (Darwin 1877; Barrett 1992; Geber *et al*. 1999). Theoretical models have motivated extensive sampling of mating-type frequencies in natural populations and

Correspondence: K. Hodgins, Fax: +1 416 978 5878; E-mail: hodgins@botany.utoronto.ca

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

have exposed complex patterns of geographical variation that commonly deviate from equilibrium expectations. Historical contingencies, especially associated with migration and founder events, have played an important role in governing these patterns (Baker 1959; Sculthorpe 1967; Barrett & Forno 1982; Eckert & Barrett 1995). However, explicit phylogeographical analysis has not been conducted on species with sexual polymorphisms until recently (Dorken & Barrett 2004).

Narcissus triandrus is a sexually polymorphic wild daffodil found in the Iberian Peninsula, which possesses considerable geographical variation in mating-type frequencies (Fernandes 1965; Barrett *et al*. 1997, 2004). The species possesses a tristylous sexual system with three floral morphs differing reciprocally in the placement of anthers and stigmas (long-, mid- and short-styled morphs, hereafter referred to as L-, M- and S-morphs). Based on large-scale surveys, three patterns are evident in the morph ratios of populations: (i) most are L-morph biased; (ii) there is a negative association between the frequencies of the L- and M-morphs; and (iii) some populations lack the M-morph and are thus dimorphic (Barrett *et al*. 2004). Throughout most of the Iberian Peninsula *N. triandrus* populations are uniformly trimorphic. However, in central Portugal dimorphic and trimorphic populations are found in close geographical proximity, and only dimorphic populations occur in the northwestern part of the range. Historical processes may have played a role in structuring geographical differentiation.

Narcissus triandrus also possesses extensive geographical variation in floral and vegetative traits. This variation has led to various taxonomic schemes, ranging from the subdivision of *N. triandrus* into several species (Pérez *et al*. 2004) to recognition of several varieties (Blanchard 1990). Here we follow Blanchard's (1990) treatment and focus our study on the two wide-ranging, largely allopatric varieties: var. *cernuus* from southern and central Spain and Portugal and var. *triandrus* from central and northern Portugal and northwestern Spain. Significantly, populations of var. *cernuus* are uniformly trimorphic, whereas var. *triandrus* is composed of trimorphic and dimorphic populations, with trimorphic populations exhibiting a wide range of M-morph frequencies compared to var. *cernuus* (Barrett *et al*. 2004). The contact zone between these varieties also corresponds to a region where there are steep clines in M-morph frequency in var. *triandrus*. Introgression of the M-morph allele from var. *cernuus* populations into dimorphic populations of var. *triandrus* could play a role in causing clines in M-morph frequency.

Previous studies investigating geographical variation in *N. triandrus* have focused on adaptive explanations associated with how floral morphology influences the mating and fertility of the style morphs (Barrett *et al*. 2004; Hodgins & Barrett 2006a, b). However, climate-driven historical processes in the Iberian Peninsula could have also influenced variation patterns, including the distribution of sexual systems. Indeed, there is a growing literature on the phylogeography of species from the Iberian Peninsula, implicating the importance of fragmentation of habitats and glacial refugia in structuring diversity (reviewed in Gómez & Lunt 2006). Concordance between neutral markers and sexual-system variation could provide evidence that historical processes have influenced the geographical differentiation of sexual systems. Alternatively, discordance between neutral markers and sexual systems could implicate current or recent selection in creating geographical patterns in sexual-system variation. Therefore, knowledge of the phylogeographical history of *N. triandrus* may provide insights into geographical variation in morphology and sexual systems in this species.

Here we use a broad geographical sampling of dimorphic and trimorphic populations of *N. triandrus* and the analysis of variation in morphological traits, chloroplast DNA (cpDNA) and nuclear microsatellite markers to address the following specific questions. What is the genetic structure of populations and does this provide information on the phylogeographical history of *N. triandrus* in the Iberian Peninsula? Are there differences between the patterns of variation in neutral genetic markers in trimorphic vs. dimorphic populations and between the two varieties of *N. triandrus*? Specifically, can these patterns inform us about the role that historical processes may have played in influencing the distribution of sexual systems? Finally, is there morphological or molecular evidence of hybridization between the varieties of *Narcissus triandrus* and, if so, could hybridization influence the frequency of the M-morph within populations?

Materials and methods

Study species

Narcissus triandrus is a widespread geophyte found abundantly in pine or oak stands on acidic soils in mountainous and hilly regions of the Iberian Peninsula. Reproductive plants flower in spring and produce a single stem per season bearing between one and nine flowers per inflorescence (mean = 1.6) that last for up to 14 days. Nectar-collecting long-tongued bees, primarily *Anthophora* and *Bombus* species, pollinate flowers. Seeds have no obvious dispersal mechanism and fall in the vicinity of maternal plants. *Narcissus triandrus* var. *cernuus* (Salisb.) Baker occurs in central and southern parts of the species' range and is characterized by a smaller stature, pale lemon flowers, and usually produces one or two flowers per plant when in bloom. In contrast, *N. triandrus* L. var. *triandrus* occurs in northern Portugal and northwestern Spain and is taller, with larger, white-to-cream flowers that are often produced in greater numbers per inflorescence.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

Sampling

We sampled 89 *N triandrus* populations throughout the range of the species during March and April of 2002–2004. Most populations were separated by at least 5–10 km and our sampling reflected the overall density of the species throughout the region. We collected leaf samples from between 10 and 286 individuals in 39 var. *cernuus* and 50 var. *triandrus* populations (18 dimorphic and 32 trimorphic). We extracted total genomic DNA from individuals using the Puregene™ DNA isolation kit (Gentra Systems). We obtained morph ratios by randomly sampling flowering stems and classifying them according to floral morph in all but four of the populations sampled for DNA, although we recorded sexual system in all populations. We collected a minimum of 100 individuals unless the population size was less than 100 and then all individuals were sampled. Detailed localities and data on floral-morph ratios are available from the first author on request.

Chloroplast DNA sequencing

We first surveyed populations for variation using universal cpDNA primer pairs corresponding to noncoding intergenic regions of cpDNA (Taberlet *et al*. 1991; Demesure *et al*. 1995; Dumolin-Lapegue *et al*. 1997). Initially we attempted the polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) method but were unable to identify adequate levels of variation. To identify more variation we sequenced SfM and seven single base-pair substitutions were detected that combined to form five haplotypes (Table 1). The original universal primers for SfM did not provide adequate sequence quality. Therefore, we designed two nested primers that we used to sequence the region: *trn*Snt-GAGATGATACACTAACTATCTGT-AGAAACATC (*trn*S) and trnfMnt-CCAACCATCAGAA-AATACAA (*trn*fMnt). We sequenced 882 bp of this region in both directions for 190 individuals (mean = 2.5 per population, range = 2–7 individuals) in 75 populations using these primers.

We performed DNA amplification using the following conditions: 50 ng of genomic DNA in a 25 μL PCR volume, along with 0.1 μM of *trn*S and *trn*fMnt, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 1.25 U Taq polymerase (Fermentas) and 1X PCR buffer with $(NH_4)_2SO_4$. The cycling conditions were 4 min initial denaturing followed by 40 cycles of: denaturing for 1 min at 94 °C; annealing at 62 °C for 1 min; and extension for 2 min at 72 °C. A final extension followed at 72 °C for 10 min. Sequencing reactions were performed with the Genome Laboratory DTCS QSK (Beckman Coulter) and run on the Beckman CEO™ 8000 Genetic Analysis System. The sequencing reactions contained 2.5 μL of enzymatically cleaned PCR product, 2.5 μL of Beckman dye terminator cycle sequencing mix, 1.5 μL of 1X PCR buffer with 1.5 mM $MgCl_2$, 3 μL water, and 0.5 μL of either trnSnt or *trn*fMnt. The cycling conditions were 96 °C for 20 s, 55 °C for 20 s, and 60 °C for 4 min, repeated 50 times. The products were cleaned by ethanol precipitation according to the manufacture's instructions. We manually edited and aligned sequences using BIOEDIT (version 5.0.9).

Analysis of chloroplast DNA sequence variation

We analysed chloroplast diversity using FSTAT (Goudet 2001) and calculated haplotype diversity for each population (H_S) and over the entire sample (H_T). To test for differences between sexual systems and taxonomic varieties in haplotype diversity we performed permutations (1000 permutations of populations between groups) using FSTAT (Goudet 2001).

We generated a haplotype network using the statistical parsimony procedure outlined by Templeton *et al*. (1992). To analyse the historical signal of the cpDNA variation, we used GEODIS (Posada *et al*. 2000), which implements the nested-clade analysis (NCA) designed by Templeton *et al*. (1995). We chose to statistically assess and interpret the

Table 1 The cpDNA sequence variation in *Narcissus triandrus*. The nucleotide position of each polymorphism is shown for the 882 bp region of the chloroplast genome that we sequenced. In addition, the haplotype as well as the taxonomic variety and sexual system associated with each haplotype are listed. GenBank accession numbers for these sequences are also listed

Haplotype (GenBank)	Variety	Sexual system	Nucleotide position								
			16	112	186	202	203	238	700	715	756
A (DQ974200)	*cernuus*	trimorphic	T	G	C	T	A	A	A	—	T
B (DQ974201)	*triandrus*	dimorphic/trimorphic	C	A	C	G	A	A	G	—	T
C (DQ974202)	*cernuus*	trimorphic	T	G	C	T	A	T	A	—	T
D (DQ974203)	*triandrus*	dimorphic/trimorphic	C	A	C	G	G	A	G	—	T
E (DQ974204)	*triandrus*	dimorphic/trimorphic	C	A	C	G	A	A	G	—	G
F (DQ974205)	*triandrus*	trimorphic	C	A	A	G	A	A	G	—	G
G (DQ974206)	*triandrus*	trimorphic	C	A	C	G	A	A	G	T	T

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

geographical pattern of the haplotypes using NCA, bearing in mind its limitations (see Knowles & Maddison 2002; Petit & Grivet 2002; Templeton 2002, 2004). We hierarchically grouped the haplotypes into clades (see Templeton *et al*. 1992 for details). We then assessed the significance of the dispersion distance of a clade, D_c, and the displacement distance, D_n, by permutation tests of the data (1000 permutations). To assess historical processes within each level of the nested cladogram we used the most recent version of the dichotomous key (http://darwin.uvigo.es/download/geodisKey_11Nov05.pdf) developed by Templeton (1998).

We performed a molecular analysis of variance (AMOVA) using ARLEQUIN 2.0 (Schneider *et al*. 2000) to examine the partitioning of genetic variation within and between populations, as well as between the two varieties and sexual systems. Estimates were obtained using haplotype frequencies and molecular information as given by the distance matrix among haplotypes (http://jan.ucc.nau.edu/~cvm/latlongdist.html). We used FSTAT (Goudet 2001) to test for differences in population differentiation between the sexual systems and varieties (F_{ST}) with 1000 permutations of populations between groups.

Microsatellite loci

We used five primers: NT26, NT63, NT113, NT154 and NT155 to assess population genetic structure at microsatellite loci in *N. triandrus* (Hodgins *et al*. 2006). In total we sampled 24 populations, focusing on portions of the range (central and northern Portugal and northwestern Spain) with the greatest variation in morph frequencies and steep transitions between trimorphic and dimorphic populations, and between taxonomic varieties. We concentrated our sampling in this region because we were interested in determining the level of population differentiation between these different groups. Our sample included 5 trimorphic populations of var. *cernuus* and 7 dimorphic and 12 trimorphic populations of var. *triandrus*. We sampled 32–115 individuals per population (mean = 38.5 individuals) for the five microsatellite loci. All but three of these populations (populations 234, 239 and 260) were also included in the cpDNA survey of sequence variation. For the three additional populations we used two PCR-RFLP markers, SfM-*Nde*I and SfM-*Nsi*I, developed from the sequenced SfM region that cut at polymorphic sites 112 and 202 (Table 1). We used variation at these sites as molecular markers for distinguishing the two taxonomic varieties. We performed DNA amplification using the above conditions for a standard PCR. DMSO was added to NT26 reactions to a final concentration of 5%. The cycling conditions were 4 min initial denaturing followed by 40–50 cycles of: denaturing for 30 s at 94 °C; annealing at 59 °C for 30 s; and extension for 30 s at 72 °C with a 72 °C 10 min final extension. For markers NT63 and NT154, PCR products were sized on the Beckman CEQ™ 8000 Genetic Analysis System using a 400-bp size standard, according to the manufacturer's instructions. Fragment sizes were estimated using the Beckman Coulter 8000 fragment analysis software. For markers NT26, NT113 and NT155 we sent the PCR reactions to the Genetic Analysis Facility of the Centre for Applied Genomics (Hospital for Sick Children, Toronto, ON). Fragment analysis and size alignments were conducted using GENEMAPPER® version 3.5 software.

Analysis of microsatellite variation

For each microsatellite locus, we assessed genetic polymorphism by calculating the observed number of alleles (n_a), the average observed heterozygosity (H_O), gene diversity (H_S) across populations, and total gene diversity (H_T). For each population we estimated genetic diversity across all loci using percentage of polymorphic loci (P), the observed number of alleles (n_a), the average observed heterozygosity (H_O), gene diversity (H_S) and allele richness (R_S). R_S is a measure of the number of alleles that is independent of sample-size (Petit *et al*. 1998). We determined levels of genetic differentiation among populations using F_{ST} (Weir & Cockerham 1984) and R_{ST}. R_{ST} is an estimator of gene differentiation that accounts for differences in allele size and assumes a stepwise-mutation model (Slatkin 1995). We assessed the appropriateness of R_{ST} vs. F_{ST} for our data using SPAGEDI (version 1.2d; Hardy & Vekemans 2002) through allele size permutation (1000 permutations). We also used MICRO-CHECKER to identify the presence of null alleles at each locus for each population (Van Oosterhout *et al*. 2004). Size homoplasy is a concern when using microsatellites and this could potentially affect population genetic parameters and historical reconstructions made in this study. Size homoplasy in microsatellites can pose problems when population sizes are large, mutation rates are high and there are strong constraints on allele size (Estoup *et al*. 2002).

We used STRUCTURE to identify the presence of population structure at microsatellite loci. STRUCTURE implements a model-based clustering method assigning individuals to populations and identifying migrants and individuals resulting from admixture (Pritchard *et al*. 2000). The number of populations (clusters), K, was set from 1 to 10. Each K was replicated 3–5 times for 1 000 000 iterations after a burn-in period of 50 000, without any prior information on the population of origin of each sampled individual. We used the admixture model, in which the fraction of ancestry from each cluster is estimated for each individual, and allowed for correlated allele frequencies. STRUCTURE identified two population clusters that generally corresponded to the two taxonomic varieties. To compare populations, we averaged data from individuals within

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

populations to estimate the fractions of ancestry. To assess the level of admixture between the two varieties we repeated the analysis, in the same manner as above. However, in the second analyses we used prior information on the population of origin of each sampled individual based on the cpDNA markers we established for identifying the varieties. We set MIG PRIOR at 0.001, 0.05 and 0.1 and the analysis was conducted in the same manner as above. We chose the number of clusters (K) based on the second order rate of change of the likelihood function with respect to K (ΔK). The modal value of the distribution of ΔK corresponds to the real K in simulations conducted by Evanno *et al*. (2005). We calculated DA genetic distances (Nei *et al*. 1983) from allele frequencies and used the resulting distance matrix to create a neighbour-joining tree. We assessed the reliability of the tree using 1000 bootstraps. This analysis was done using DISPAN (Ota 1993).

We analysed patterns of diversity with FSTAT (Goudet 2001) by calculating global estimates of R_S, H_O, H_S, F_{IS} and F_{ST} and tested for differences between groups using 10 000 permutations where the units of randomization were populations. The groups were based on sexual system, taxonomic variety, or the groups identified by structure. We performed these analyses using all five microsatellite loci. We also examined the extent of genetic subdivision among groups using analysis of molecular genetic variation (AMOVA) with F_{ST} (Excoffier *et al*. 1992) as implemented in ARLEQUIN 2.0 (Schneider *et al*. 2000).

To detect if there was significant isolation-by-distance, we performed Mantel tests using ARLEQUIN 2.0 (Schneider *et al*. 2000). We calculated geographical distances between pairs of populations as above. We used 10 000 random permutations between the matrix of pairwise genetic differentiation between populations $F_{ST}/(1 - F_{ST})$, and the matrix of the natural logarithm of geographical distance. We performed the analyses on the whole data set and separately for each cluster (see below). We also conducted partial Mantel tests with matrices of pairwise differences of morph frequency. Matrices were created for each morph, where the difference in morph frequency for each pair of populations was calculated.

We also estimated the pollen-to-seed migration ratio, $r = m_p/m_s$ (Ennos 1994) following Petit *et al*. (2005) using G_{STm} and G_{STb}, which correspond to the estimate of subdivision at maternally inherited markers and at nuclear (biparentally) inherited markers, and F_{IS}, which is the heterozygote deficit estimated with nuclear codominant microsatellite markers.

Morphological sampling and analysis

To investigate evidence for hybridization between vars. *cernuus* and *triandrus* we measured 20 randomly sampled individuals per population for 14 morphological characters in 45 populations from throughout the range (25 var. *triandrus*, 20 var. *cernuus*). The following traits were measured: plant height, leaf length, stem diameter, bulb diameter, flower size, corona length, tube length, corona width at the mouth, corona width at the maximum circumference, floral tube width at the junction of the tube and the corona, tube width at the base and tepal length using a ruler or digital callipers as described in Barrett *et al*. (2004). We also recorded leaf and flower number for each individual. Population means were taken for each variable. We used PCR-RFLP markers (SfM-*Nde*I and SfM-*Nsi*I) to identify the varieties for populations not included in the survey of the cpDNA sequence variation.

We performed a discriminate functions analysis to determine if the cpDNA molecular markers distinguishing the two varieties were congruent with the morphological variation. Discordance between the two classes of data would provide evidence for hybridization between vars. *cernuus* and *triandrus*. We carried out all statistical analysis of the morphological data using SAS version 9.1 (SAS Institute Inc.). We log transformed or square root transformed data if it exhibited non-normality and tested for within-class multivariate normality using the MULTNORM macro provided by SAS. We used stepwise discriminant analysis implemented by the STEPDISC procedure (SAS) to select a subset of variables for use in discriminating among the classes. We chose stepwise selection (backwards) for the analysis, with no variables in the model. A value of $P = 0.15$ was used for the stop criterion. We used the selected variables for analysis using the DISCRIM procedure (SAS) with the crossvalidate option. This method classifies each observation in the data set using a discriminate function computed from the other observations in the data set, excluding the observation being classified.

Results

Chloroplast DNA diversity and population structure

Using cpDNA sequence variation we identified seven haplotypes (A–G) that were linked by nine mutational steps (see Table 1; Fig. 1b). All of the mutations were single base-pair substitutions except one indel separating haplotypes B and G. Haplotypes were not equally frequent and their distribution indicated strong geographical structuring (Fig. 1a). Two of the seven haplotypes were only observed in var. *cernuus* (A and C), while the other five were only found in var. *triandrus* (B, D, E, F, and G). The haplotype network identified significant molecular subdivision between the varieties with four sequence polymorphisms separating the interior haplotypes A and B. Haplotype A was the most common and widespread haplotype (38.4%) and was found across the entire southern portion of the range. Haplotype C, which was at low frequency (3.2%), was clustered with haplotype A in the central region of Spain.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

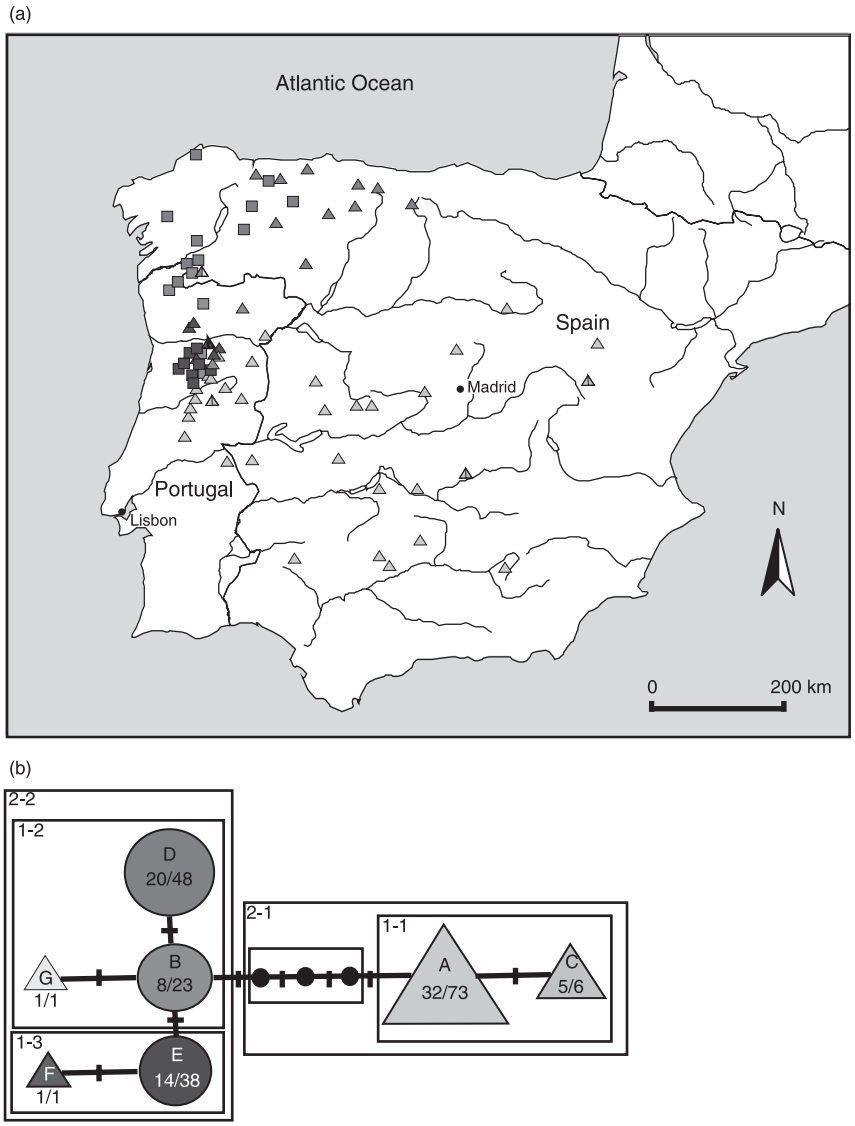

Fig. 1 Chloroplast haplotypes of *Narcissus triandrus*. (a) The geographical distribution of cpDNA haplotypes in *Narcissus triandrus*. Chloroplast DNA (cpDNA) haplotypes are represented by different colours and polymorphic population are represented with both colours. Sample size was 2–7 individuals per populations and 75 populations (triangles = trimorphic populations and squares = dimorphic). (b) A network of cpDNA haplotypes. Each line represents a single mutation, and small filled circles indicate haplotypes not present in the sample. The boxes indicate one-, two- and three-step clades which are labelled with the first number corresponding to the nesting level, and the second number representing groups at the same nesting level. The shapes representing each haplotype reflect the sexual system of populations associated with each haplotype (triangle = trimorphic and circle = dimorphic and trimorphic). The sizes of the shapes are proportional to the frequency of each haplotype. The numbers indicate the number of populations and individuals in which each haplotype was found.

Central Portugal contained the greatest haplotype variation with haplotypes B, E, F and G. The interior haplotype B (12.0%) was found in central and northern Portugal. Haplotype E (20.0%) was only found in central Portugal, while haplotype D (25.3%) was found exclusively in northwestern Spain and northern Portugal. Of the seven haplotypes, two were only represented in single individuals (G and H; both 0.5%). Chloroplast diversity within populations was rare, with only 6 of the 75 populations (two var. *triandrus* and four var. *cernuus* populations) containing more that one haplotype. All three common haplotypes associated with var. *triandrus* occurred in both dimorphic and trimorphic populations. Haplotypes A and C were only found in trimorphic populations because all var. *cernuus* populations are trimorphic.

Haplotype diversity within populations was low (mean $H_S = 0.068$) although there was a higher level of total diversity overall ($H_T = 0.66$). Var. *cernuus* had lower haplotype diversity ($H_T = 0.17$) compared to var. *triandrus* ($H_T = 0.64$) but the taxonomic varieties had similar amounts of haplotype diversity when averaged across populations (var. *cernuus* $H_S = 0.09$, var. *triandrus* $H_S = 0.03$, $P = 0.32$). Overall, the diversity of dimorphic and trimorphic populations was similar (dimorphic $H_T = 0.64$, trimorphic var. *triandrus* $H_T = 0.62$, trimorphic all populations $H_T = 0.61$). When comparing sexual systems, dimorphic populations had no within population diversity ($H_S = 0$) although dimorphic population diversity did not differ from var. *triandrus* trimorphic populations ($H_S = 0.07$, $P = 0.25$) or across all trimorphic populations ($H_S = 0.08$, $P = 0.14$).

Among all populations genetic differentiation was high ($F_{ST} = 0.92$, $G_{ST} = 0.90$), and taking molecular distance into account this increased the estimate of population differentiation ($F_{ST} = 0.98$). Var. *cernuus* populations had significantly

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

Table 2 The results of a nested clade analysis of cpDNA variation in *Narcissus triandrus*. See methods for further details

Clade	χ^2-statistic	Probability	Inference chain	Inferred pattern
1-1	50.50	< 0.05	1-2-11-17	Inconclusive outcome
1-2	83.89	< 0.001	1-2-3-5-15-21	Gradual range expansion followed by fragmentation
1-3	18.98	0.45	na	Panmixia or small sample size
2-2	111.00	< 0.001	1-2-3-4	Restricted gene flow followed by isolation-by-distance
Total cladogram	189.00	< 0.001	1-19	Allopatric fragmentation

The probability refers to the frequency with which the 1000 randomly generated chi-square statistics were equal to or greater than the observed chi-square. Haplotype groups without genetic and/or geographical variation are not listed.

lower levels of population differentiation than var. *triandrus* (var. *cernuus* $F_{ST} = 0.40$, var. *triandrus* $F_{ST} = 0.94$, $P < 0.01$). Dimorphic populations had significantly greater population differentiation when compared to all trimorphic populations (dimorphic $F_{ST} = 1$, trimorphic $F_{ST} = 0.87$, $P = 0.05$) but not when compared to var. *triandrus* trimorphic populations ($F_{ST} = 0.90$, $P = 0.23$).

The nested clade analysis (NCA) identified significant genetic structure for the 1-1, 1-2, 2-2 clades and the entire cladogram (Table 2). Allopatric fragmentation between clade 2-1 and 2-2 was inferred indicating that the two varieties likely differentiated in geographical isolation. Restricted gene flow and isolation-by-distance were inferred for clade 2-2 while gradual range expansion followed by fragmentation was inferred for clade 1-2. Clades 1-1 showed significant geographical structuring of haplotypes but no significant D_n or D_c values were identified.

AMOVA of all populations demonstrated that most of the molecular variation was distributed among populations (94.95%). Classifying populations as either vars. *cernuus* or *triandrus* explained the largest portion of the variation (86.82%; Table 3a). Among all populations sexual system explained (42.13%) of the variation (Table 3b); however, within var. *triandrus* sexual system did not explain any of the molecular variation (–4.64%; Table 3c) indicating that the varietal difference is the principal cause of haplotype differentiation and not sexual system.

Table 3 Analysis of molecular variance (AMOVA) for *Narcissus triandrus* cpDNA sequence variation between (a) populations of *Narcissus triandrus* vars. *cernuus* and *triandrus* (b) dimorphic and trimorphic populations of *Narcissus triandrus* (c) dimorphic populations and trimorphic populations of *Narcissus triandrus* var. *triandrus*. * denotes significant values $P < 0.01$

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
(a)				
Among groups	1	199.21	2.15	86.82*
Among populations within groups	73	56.77	0.30	12.02*
Within populations	115	3.30	0.03	1.16*
Total	189	259.28	2.47	
(b)				
Among groups	1	65.85	0.77	42.13*
Among populations within groups	73	190.13	1.02	56.29*
Within populations	115	3.30	0.02	1.58*
Total	189	259.38	1.82	
(c)				
Among groups	1	0.33	–0.02	–4.64
Among populations within groups	40	55.81	0.56	100.82*
Within populations	69	1.30	0.02	3.82*
Total	110	54.52	0.49	

Microsatellite diversity and population structure

Diversity estimates varied among microsatellite loci (Table 4) and among populations (Table 5). The total number of alleles per locus in our sample of 923 individuals ranged from 11 to 35 with a total of 106 alleles at the five loci. Average gene diversity (H_S) among the loci was 0.57, range 0.21–0.87. Gene diversity (H_T) estimates among all populations ranged from 0.27 to 0.91 with an average of 0.62 ± 0.11. The level of polymorphism within populations ranged from 80 to 100%. Values for gene diversity (H_S) ranged from 0.41 to 0.77, observed heterozygosity (H_O) 0.32–0.76, and allelic richness (R_S) 5.18–8.5. The F_{IS} values for all populations ranged from 0.0 to 0.38 with an average of 0.14 ± 0.07 indicating relatively low levels of inbreeding within populations, although half of the 24 populations had F_{IS} values significantly greater than 0. Previous estimates of inbreeding using allozyme markers were found to be slightly lower ($F_{IS} = 0.06$; Hodgins & Barrett 2006b). This discrepancy could be, in part, due to null alleles. We detected the presence of null alleles in some populations at all five loci, although in the majority of populations we did not detect significant frequencies of null alleles. The mean frequencies among populations for each locus using the Brookfield (1996) estimator were –0.02 ± 0.05 for NT155, 0.03 ± 0.01 for NT113, 0.14 ± 0.16 for NT26, 0.04 ± 0.02 for NT63, and 0.003 ± 0.008 for NT154.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

Table 4 A comparison of genetic diversity at five microsatellite loci in *Narcissus triandrus*: observed allele number (n_a); observed heterozygosity (H_O); gene diversity (H_S); overall gene diversity (H_T); among population genetic differentiation (F_{ST}); gene differentiation accounting for variance in allele size (R_{ST})

Locus	n_a	H_O	H_S	H_T	F_{ST}	R_{ST}
NT155	16	0.25	0.21	0.27	0.22	0.26
NT26	25	0.58	0.66	0.72	0.09	0.03
NT113	35	0.58	0.87	0.91	0.06	0.10
NT63	19	0.56	0.63	0.74	0.17	0.42
NT154	11	0.44	0.45	0.47	0.04	0.02
Mean ± SE	21.2 ± 4.1	0.48 ± 0.06	0.57 ± 0.11	0.62 ± 0.11	0.1 ± 0.03	0.17 ± 0.07

We calculated all parameters using FSTAT software. F_{ST} values were estimated according to Weir & Cockerham (1984). Overall loci, R_{ST} was determined using the Goodman (1997) estimator.

Table 5 Descriptive statistics of genetic variability for 24 populations of *Narcissus triandrus*: Percent of polymorphic loci (P); observed allele number (n_a); observed heterozygosity (H_O); Nei's unbiased gene diversity (H_S); allelic richness calculated with the rarefaction method (R_S); inbreeding coefficient (F_{IS})

Population	P	n_a	H_O	H_S	R_S	F_{IS}
Trimorphic populations var. *cernuus*						
142	100	8.6	0.54	0.68	8.13	0.20*
147	100	8.8	0.61	0.70	8.26	0.13*
148	100	9.0	0.73	0.77	8.50	0.05
173	100	7.8	0.61	0.68	7.38	0.10
184	100	8.8	0.69	0.74	8.24	0.07
Trimorphic populations var. *triandrus*						
153	100	8.4	0.76	0.71	7.89	0.00
154	100	7.4	0.47	0.51	6.74	0.07
156	80	5.0	0.48	0.46	4.83	0.00
157	100	5.0	0.28	0.41	4.79	0.33*
207	100	9.6	0.67	0.72	8.58	0.08
209	80	5.4	0.39	0.48	5.18	0.20
222	80	6.6	0.40	0.51	6.22	0.22*
234	100	5.4	0.36	0.49	5.19	0.27*
236	80	5.0	0.32	0.43	4.77	0.25*
239	80	5.6	0.33	0.42	5.21	0.21*
254	100	8.0	0.48	0.48	5.67	0.00
260	100	9.2	0.68	0.74	8.47	0.08
Dimorphic populations var. *triandrus*						
152	100	7.4	0.43	0.56	7.00	0.23*
163	80	7.6	0.42	0.59	7.16	0.28*
204	80	9.0	0.49	0.51	5.92	0.03
223	80	7.2	0.34	0.55	6.83	0.38*
231	100	7.4	0.35	0.45	6.77	0.22*
238	80	7.4	0.30	0.45	6.80	0.34*
246	80	6.4	0.42	0.50	5.91	0.17
Mean ± SE	91.7 ± 2.1	7.3 ± 0.30	0.48 ± 0.03	0.57 ± 0.02	6.69 ± 0.27	0.14 ± 0.07

We calculated all parameters with FSTAT software. *F_{IS} values significantly different from 0 based on 10 000 randomizations with table wide $P < 0.05$.

Two main clusters of populations were identified using STRUCTURE; a northwest cluster, comprised of var. *triandrus* populations, and a south-central cluster, composed of var. *cernuus* populations and three var. *triandrus* populations. These three var. *triandrus* populations were population 153, 207 and 260, where the average fraction of ancestry for the south-central group was > 86% for all three populations. However, many populations showed some evidence of mixed ancestry (Fig. 2a). To assess evidence for hybridization between the two varieties we reanalysed the data to

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

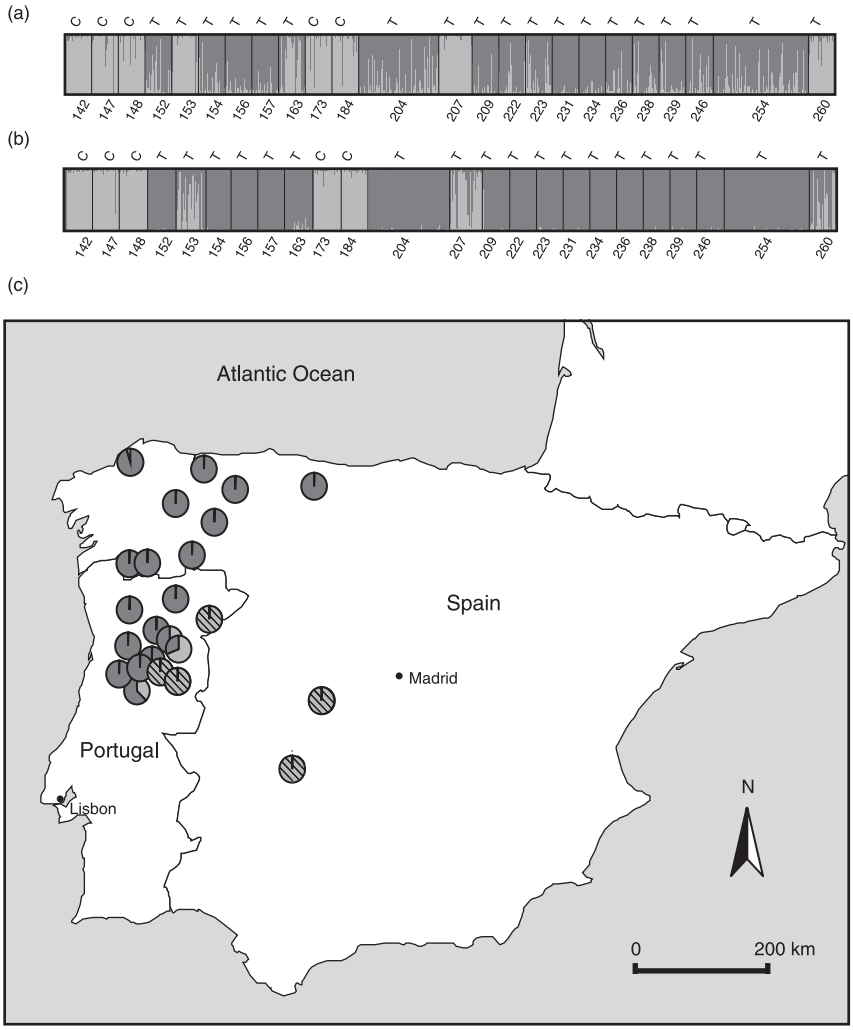

Fig. 2 The population structure of *Narcissus triandrus* based on five polymorphic microsatellite loci. Between 32 and 115 individuals were sampled from 24 populations. (a) The two clusters detected from STUCTURE analysis with no prior population knowledge. (b) The results of STRUCTURE using prior population knowledge based on cpDNA markers for each variety. Population number and variety as determined by the cpDNA markers (C = *Narcissus triandrus* var. *cernuus*, T = var. *triandrus*) are shown (c) The geographical distribution of the two clusters from the results of STRUCTURE using prior population knowledge based on cpDNA markers for each variety. Pie charts represent the proportion of each of the clusters in each population and cross hatching represents populations containing *Narcissus triandrus* var. *cernuus* cpDNA markers.

examine the degree of admixture using the prior population grouping based on cpDNA markers (Fig. 2b, c). For these same three populations of var. *triandrus*, the average fraction of ancestry for the south-central group was high (population 153, 48%; population 207, 69%; and population 260, 37%), revealing incongruence between the cpDNA and microsatellite markers.

The neighbour-joining tree reflected the groupings identified by STRUCTURE with all var. *cernuus* populations grouped together along with the three populations with var. *triandrus* cpDNA markers (populations 153, 207 and 260; Fig. 3). All of these populations were trimorphic. All other populations of var. *triandrus* formed a separate group with dimorphic and trimorphic populations interspersed among the remaining portions of the tree.

Genetic differentiation was examined using both the infinite alleles and stepwise-mutation models. Mean F_{ST} and R_{ST} were 0.10 ± 0.03 and 0.17 ± 0.07, respectively. SPAGEDI revealed that R_{ST} was not significantly different than pR_{ST}, which represents R_{ST} after allele size permutations (across all loci and populations $P = 0.148$). However, for two loci R_{ST} differed significantly from pR_{ST} (NT26 $P < 0.05$; NT63 $P < 0.001$). Measures of genetic diversity differed significantly between the two groups identified by STRUCTURE (northwest $R_S = 5.94$, south-central $R_S = 8.18$, $P < 0.001$; northwest $H_O = 0.41$, south-central $H_O = 0.66$, $P < 0.001$; northwest $H_S = 0.49$, south-central $H_S = 0.72$, $P < 0.001$), although levels of inbreeding (F_{IS}) and population differentiation F_{ST} were not significantly different. Genetic diversity was greater in var. *cernuus* compared to var. *triandrus* (var. *triandrus* $R_S = 6.31$, var. *cernuus* $R_S = 8.10$, $P < 0.01$; var. *triandrus* $H_O = 0.45$, var. *cernuus* $H_O = 0.64$, $P < 0.01$; var. *triandrus* $H_S = 0.52$, var. *cernuus* $H_S = 0.71$, $P < 0.01$) and population differentiation was lower in var. *cernuus* (var. *cernuus* $F_{ST} = 0.03$, var. *triandrus* $F_{ST} = 0.09$, $P < 0.05$), although, levels of inbreeding (F_{IS}) between the varieties did not differ. Measures of genetic diversity were not significantly different between dimorphic

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

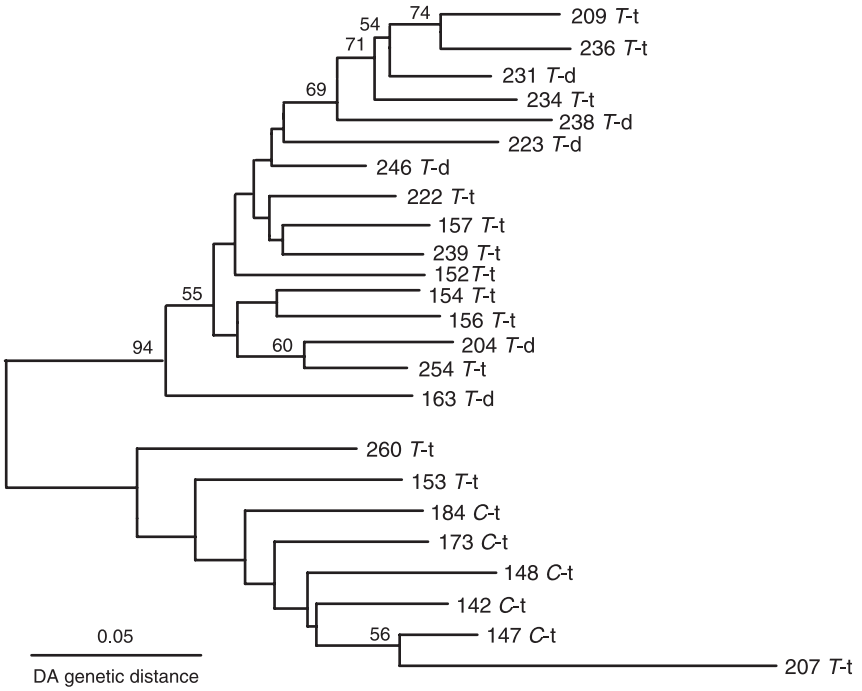

Fig. 3 A neighbour-joining tree of 24 dimorphic and trimorphic populations of *Narcissus triandrus* based on five microsatellite loci. DA genetic distances were calculated from allele frequencies using DISPAN. Population number, variety as determined by the cpDNA markers (*C* = *Narcissus triandrus* var. *cernuus*, *T* = var. *triandrus*) and sexual system (d = dimorphic and t = trimorphic) are indicated.

and trimorphic populations, although genetic differentiation among populations was significant, both within var. *triandrus* and among all populations (dimorphic $F_{ST} = 0.06$, all populations trimorphic $F_{ST} = 0.11$, $P < 0.01$; var. *triandrus* trimorphic $F_{ST} = 0.09$, $P < 0.01$). The pollen and seed migration ratio was estimated at 98.5 indicating that the contribution of pollen to gene flow is almost 100-times greater than the contribution of seeds in this species.

AMOVA of all populations demonstrated that most of the molecular variation was distributed within populations (90.08%) rather than among populations (9.92%). AMOVA confirmed significant differentiation between the northwest and south-central clusters (10.75%; Table 6a). Taxonomic variety also explained a significant percentage of the variation (7.65%; Table 6b). Among all populations, sexual system explained 0.72% of the variation (Table 6c) and within var. *triandrus* sexual system explained 0% of the molecular variation (Table 6d).

Mantel tests among all populations did not identify significant isolation-by-distance ($r = 0.12$, $P = 0.10$). Within the two groups identified by STRUCTURE, there was significant isolation-by-distance in the northwest group ($r = 0.19$, $P < 0.05$) but not in the south-central group ($r = -0.07$, $P = 0.61$). Within dimorphic populations there was no evidence of isolation-by-distance ($r = 0.03$, $P = 0.42$), nor was there among all trimorphic populations ($r = 0.12$, $P = 0.10$). Differences in the frequency of the M-morph were significantly associated with genetic distance among all populations ($r = 0.54$, $P < 0.001$; Fig. 4) and within the northwest group ($r = 0.29$, $P < 0.05$), but not within the south-central group ($r = 0.12$, $P = 0.30$). Partial Mantel tests identified significant

Table 6 Analysis of molecular variance (AMOVA) for *Narcissus triandrus* using five microsatellite loci between (a) northwest and south-central groups identified by STRUCTURE (b) *Narcissus triandrus* vars. *cernuus* and *triandrus* (c) dimorphic and trimorphic populations of *Narcissus triandrus* (d) dimorphic and trimorphic populations of *Narcissus triandrus* var. *triandrus*. * denotes significant values $P < 0.01$

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
(a)				
Between groups	1	129.15	0.16	10.75*
Between populations within groups	22	146.49	0.07	4.70*
Within populations	1822	2324.37	1.28	84.55*
Total	1845	2600.01	1.51	
(b)				
Between groups	1	69.45	0.11	7.65*
Between populations within groups	22	206.19	0.11	7.10*
Within populations	1822	2324.37	1.28	85.25*
Total	1845	2600.01	1.49	
(c)				
Between groups	1	23.44	0.01	0.72
Between populations within groups	22	252.20	0.14	9.58*
Within populations	1822	2324.37	1.28	89.70*
Total	1845	2600.01	1.42	
(d)				
Between groups	1	8.66	0.00	0.00
Between populations within groups	17	177.23	0.12	9.06*
Within populations	1507	1815.26	1.20	91.49*
Total	1525	2001.14	1.32	

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

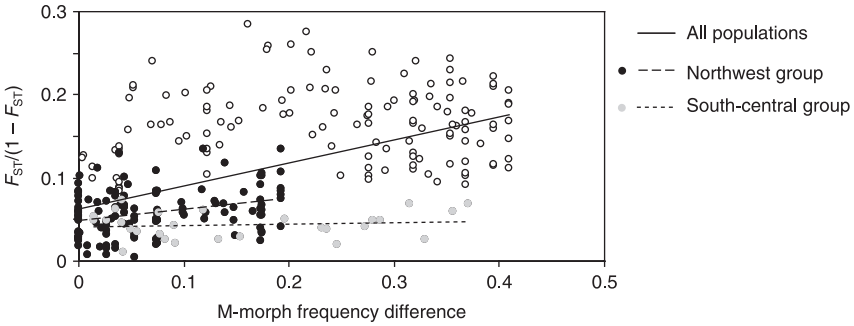

Fig. 4 The correlation between pairwise genetic distances and pairwise M-morph frequency differences among 24 populations of *Narcissus triandrus*, using variation at five microsatellite loci. M-morph frequency distance represents the difference in M-morph frequency between pairs of populations. Open circles represent comparisons between northwest and south-central populations; black circles represent comparisons within the northwest group; and grey circles represent comparisons within the south-central group (all populations $r = 0.54$, $P < 0.001$; northwest $r = 0.31$, $P < 0.05$; south-central $r = 0.13$, $P = 0.31$).

associations between differences in M-morph frequency ($r = 0.54$, $P < 0.001$) but not geographical distance among all populations ($r = 0.05$, $P = 0.24$), although both geographical distance ($r = 0.23$, $P < 0.05$) and M-morph difference ($r = 0.31$, $P < 0.05$) were significant in the northwest group. In contrast, neither were significant in the south-central group (geographical $r = –0.072$, $P = 0.60$; M-morph $r = 0.13$, $P = 0.31$). Partial Mantel tests also revealed a significant correlation between genetic distance and L-morph frequency differences (geographical $r = 0.76$, $P = 0.18$; L-morph $r = 0.30$, $P < 0.01$) among all populations. In contrast, S-morph frequency difference was marginally correlated with genetic distance, but only in the south-central group (geographical $r = –0.06$, $P = 0.57$; S-morph difference $r = 0.47$, $P = 0.056$).

Morphology

Of the 13 morphological traits measured, the STEPDISC procedure identified five variables that distinguished the two classes of populations corresponding to taxonomic varieties (Table 7). Using these variables the discriminant analysis misclassified five populations (148, 153, 207, 260, 270), including the three that were identified by microsatellite analysis (Fig. 3). All of these populations were located along contact zones between the two varieties. Population 148 was grouped with var. *triandrus* and the remaining populations clustered with var. *cernuus*.

Table 7 Means ± SE of floral and vegetative traits of the two varieties of *Narcissus triandrus*. The results of a one-way ANOVA which compared each trait are also given (GLM procedure, SAS). Variables identified by STEPDISC for use in the discriminant functions analysis are in bold. All measurements are in mm except plant height and leaf length (cm)

Variable	var. *cernuus*	var. *triandrus*	$F_{1,43}$	P
Plant height	19.76 ± 0.73	22.12 ± 1.09	2.64	0.11
Bulb diameter	12.4 ± 0.22	12.67 ± 0.26	0.40	0.53
Stem width	2.36 ± 0.07	2.78 ± 0.11	8.95	< 0.01
Leaf length	21.92 ± 1.06	22.26 ± 1.17	0.01	0.94
Flower length	24.80 ± 0.37	26.75 ± 0.58	7.12	< 0.05
Corona length	9.40 ± 0.20	10.79 ± 0.36	10.67	< 0.01
Tube length	15.42 ± 0.18	16.06 ± 0.28	3.07	0.087
Tube width at base	1.50 ± 0.03	1.80 ± 0.04	27.41	< 0.001
Tube width at junction	5.62 ± 0.08	5.38 ± 0.08	4.54	< 0.05
Corona width at mouth	8.26 ± 0.18	9.22 ± 0.22	10.21	< 0.01
Corona maximum width	9.59 ± 0.18	9.89 ± 0.20	1.17	0.28
Tepal length	15.53 ± 0.27	20.17 ± 0.73	28.98	< 0.001
Flower number	1.35 ± 0.01	1.61 ± 0.08	6.86	< 0.05
Leaf number	1.15 ± 0.04	1.84 ± 0.09	46.25	< 0.001

Discussion

Our investigation of population genetic structure in *N. triandrus* identified three main findings. First, cpDNA variation exhibited strong geographical differentiation associated with the morphological differences that distinguish the two main taxonomic varieties. Second, discordance between cpDNA, microsatellite and morphological variation indicated asymmetrical introgression between the varieties at contact zones. Finally, within var. *triandrus*, chloroplast haplotypes were not associated with sexual system, indicating that transitions between trimorphism and dimorphism are likely to have occurred within each maternal lineage. We discuss these findings in the context of historical processes associated with climatic variability of the Iberian Peninsula and contemporary microevolutionary forces.

Population structure and phylogeographical history

Studies of the postglacial migratory history of the European biota have consistently identified three primary European

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

refugia: the Iberian Peninsula, Italy and the Balkans (Hewitt 2000). For example, climatic and palynological evidence was used to support a single refugium within the southern portion of the Iberian Peninsula (Huntley & Birks 1983; Brewer *et al*. 2002). However, as the number of phylogeographical studies have accumulated the southern refugia paradigm has failed for numerous species and patterns of re-colonization have been found to be considerably more complex. In the Iberian Peninsula, recent paleoecological data indicates the persistence of plant species in the northern portions of the region (Olalde *et al*. 2002) and phylogeographical studies have demonstrated geographical differentiation among populations of diverse species, implicating multiple refugia within this region, including northern refugia (e.g. Comes & Abbott 2000; Olalde *et al*. 2002; Paulo *et al*. 2002; Martinez-Solano *et al*. 2006). There is some discordance in the locations of refugia among species, but this is likely to reflect particular habitat requirements of the groups investigated. Several refugia coincide with areas of high endemism (Gómez & Lunt 2006) and these tend to occur in mountainous areas of the Iberian Peninsula occupied by *N. triandrus*.

Our results provide evidence that the complex topography and extensive climatic variation of the Iberian Peninsula has promoted isolation among populations of *N. triandrus* (Fig. 1a). The NCA of cpDNA variation supports the hypothesis of past fragmentation of *N. triandrus* into populations occupying two distinct refugia. The nonoverlapping ranges and large number of mutational steps differentiating clades 2-1 and 2-2 suggest a division of *N. triandrus* into northwest and south-central refugia that likely persisted throughout the Quaternary. Clades 2-1 and 2-2 are concordant with previously described taxonomic groupings based on morphology (Blanchard 1990).

Our discriminant functions analysis (Table 7) indicated that morphological differentiation corresponded to divergence at the chloroplast locus throughout most of the geographical range of *N. triandrus*. The results of the AMOVA of cpDNA variation revealed significant molecular divergence between the varieties (86.82%). In addition, microsatellite variation generally supported the same geographical subdivision as the cpDNA haplotype data. AMOVA of the microsatellite variation identified a significant percent of the variation (7.65%) occurring between the two varieties, as defined by the cpDNA markers. The neighbour-joining tree and analysis of 24 populations using STRUCTURE also identified a northwest and south-central group that was generally congruent with varietal differences (Figs 2 and 3).

A phylogeographical signal was also evident in the cpDNA haplotype variation at lower clade levels. The haplotype distribution of the common var. *triandrus* haplotypes, B, D and E, suggest more recent east–west barriers to gene flow (Fig. 1a). The NCA of the 2-2 clade identified restricted gene flow and isolation-by-distance as historical processes influencing the distribution of clades 1-2 (composed of haplotype B, D and G) and 1-3 (composed of haplotypes E and F). The NCA of the 1-2 clade identified gradual range expansion followed by fragmentation between interior haplotype B and tip haplotype D. Haplotype D, located exclusively in the northwestern corner of the Iberian Peninsula, is found in the western half of the Cantabrian-Atlantic Mountains and the subcoastal mountains of Galicia. The Picos de Europa, an area of high endemism, and the warmer coastal areas of this region have been proposed as refugia for other plant species (Gómez & Lunt 2006). The ancestral haplotype B was found south of the Mino River and north of the Duero River, although some populations with this haplotype were found in central Portugal interspersed among haplotype E. Central Portugal likely acted as another refugium for *N. triandrus* and indeed the Serra de Estrela has also been implicated as a refugial area and biogeographical boundary for other Iberian species (Gómez & Lunt 2006).

The microsatellite markers identified significant but weaker geographical structuring in the northwestern portion of the range which contrasts with the stronger pattern found with cpDNA markers. The absence of strong geographical structuring in nuclear markers in this region could have at least two explanations. First, the higher mutation rates of microsatellites may result in genetic patterns that are more difficult to detect, particularly those of the distant past (Epperson 2002). Second, the observed higher gene flow among populations via pollen compared to seeds in *N. triandrus* could reduce the level of divergence among populations at nuclear loci compared to chloroplast loci. However, the pattern in the nuclear markers is still consistent with restricted gene flow and IBD in var. *triandrus*.

Contrasting patterns of molecular diversity

Comparisons of chloroplast and microsatellite variation between the two varieties of *N. triandrus* revealed contrasting patterns of diversity. This finding likely reflects differences in demography and contemporary vs. historic patterns of gene flow among populations. Variety *cernuus* had lower levels of cpDNA variation compared to var. *triandrus*, although within population diversity did not differ between the varieties. Variety *cernuus* also had significantly lower population differentiation compared to var. *triandrus*. These differences among populations of var. *triandrus* could result from fragmentation and restricted gene flow, as identified by the NCA. Admixtures of haplotypes originating from different refugia can explain high levels of haplotype diversity in other species (Petit *et al*. 2003) and may account for the higher levels of haplotype diversity in var. *triandrus*. A recent population expansion of var. *cernuus* throughout southern and central Spain from a single refugium could explain the lower gene diversity of cpDNA in this variety, although the results of the NCA were equivocal.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

In contrast, microsatellite diversity was significantly higher in var. *cernuus* compared to var. *triandrus*, as measured by allelic richness, observed heterozygosity and gene diversity. Similarly, the northern group of populations identified by STRUCTURE, composed of var. *triandrus* populations, had lower levels of diversity compared to south-central populations. Higher gene diversity in var. *cernuus* may reflect demographical differences between the varieties. If var. *cernuus* underwent a recent population expansion from a single refugia, as our cpDNA data suggests, the variation at microsatellite loci would recover more quickly due to higher mutation rates compared with the cpDNA locus. Moreover, larger census population sizes and higher plant densities that are characteristic of contemporary populations of var. *cernuus* compared to var. *triandrus* (Hodgins & Barrett 2006a and unpublished data) probably result in the maintenance of more neutral variation within var. *cernuus* populations. Other explanations for the contrasting patterns between cpDNA and nuclear microsatellites in the two varieties, such as introgression of cpDNA haplotypes from other *Narcissus* species, selective sweeps, or differences in mutation rates between the subspecies seem less likely.

As expected, the cpDNA and microsatellite data on genetic diversity and population structure in *N. triandrus* gave contrasting results (see Ennos 1994). Most of the cpDNA variation was distributed among populations (94.95%) and levels of within population diversity were very low (mean $H_S = 0.068$). As a result, genetic differentiation among populations was high ($G_{ST} = 0.90$). In contrast, microsatellite data revealed that most of the variation was within populations (90.08%) and, accordingly, estimates of gene diversity within populations were generally high (mean $H_S = 0.57$) and differentiation among populations was low ($F_{ST} = 0.10$). Low levels of genetic differentiation are commonly reported in outcrossing species using biparentally inherited markers such as allozymes (reviewed in Hamrick & Godt 1989; Hamrick & Godt 1996). These differences in diversity and population structure between maternally inherited cpDNA markers and biparentally inherited simple sequence repeat markers are a general feature of many plant species (Petit *et al*. 2005) and largely reflect the smaller effective population size of the chloroplast genome (Schaal *et al*. 1998; Zhang & Hewitt 2003). In addition, *Narcissus triandrus* seeds have no obvious mechanism for long-distance dispersal and, not surprisingly, we estimated considerably higher gene flow among populations via bee-mediated pollen dispersal than by seeds.

Local introgression between taxonomic varieties

Our analyses revealed discordance between the patterns of cpDNA variation and morphology in several populations of *N. triandrus*. This provides evidence for hybridization between the two taxonomic varieties. Incomplete lineage sorting can also produce a mismatch of genetic and morphological divergence. However, the locations of the misclassified populations along contact zones between vars. *cernuus* and *triandrus* make this hypothesis rather unlikely. Four of the five misclassified populations possessed var. *triandrus* cpDNA markers and var. *cernuus* morphology. This suggests asymmetrical gene flow with introgression of the var. *triandrus* chloroplast genome into a largely var. *cernuus* genetic background. However, morphological divergence may not always reflect genetic differentiation and group relationships based on morphological divergence alone can be misleading (Schaal *et al*. 1998). Nevertheless, throughout much of the range of *N. triandrus* there is strong concordance between morphological and genetic differentiation indicating that morphology is generally a good indicator of varietal status.

Comparisons of cpDNA and microsatellite variation revealed similar patterns of introgression in *N. triandrus*. Three populations with var. *triandrus* cpDNA markers grouped with var. *cernuus* using both STRUCTURE and the neighbour-joining tree. These three populations were also among the five in which there was a mismatch between morphology and cpDNA markers, suggesting that chloroplast capture has occurred in these populations (Rieseberg *et al*. 1990; Terauchi *et al*. 1991). Introgression of nuclear genes from var. *cernuus* into var. *triandrus*, including those that govern the style-length alleles, may contribute to the considerable heterogeneity in morph frequencies that characterize contact zones between the two varieties (Barrett *et al*. 2004). Gene flow of this type could restore missing alleles that govern the M-morph and convert dimorphic population of var. *triandrus* into populations containing all three style morphs.

Population structure, morph ratios and transitions between sexual systems

Determining the evolutionary relationships between trimorphism and dimorphism in *N. triandrus* could provide insight into the evolution, maintenance and breakdown of tristyly. Dimorphic populations are considered a likely intermediate stage in the evolution of tristyly from a monomorphic condition (Charlesworth 1979). Alternatively, stylar dimorphism could be a secondary reversion from trimorphism (Barrett & Hodgins 2006). Intraspecific gene genealogies could inform us about the evolutionary origins of the two sexual systems in *N. triandrus*. For example, if dimorphism is derived and gene flow between sexual systems is low, haplotypes found in dimorphic populations should represent a restricted subset of those in trimorphic populations.

The haplotype network and distribution of haplotypes among populations of *N. triandrus* provided little information

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

on the relationships between trimorphism and dimorphism. The ancestral haplotype A was only found in trimorphic populations, while dimorphic populations contained a restricted subset of haplotypes (three out of seven). However, within var. *triandrus* the interior haplotype B was evenly represented among dimorphic and trimorphic populations as were the other two common haplotypes E and D. Therefore, the ancestral condition cannot be inferred from these data.

Trimorphic and dimorphic populations were represented among all of the common haplotypes detected in *N. triandrus* var. *triandrus*. This suggests that the M-morph could have been lost and/or gained within each maternal lineage. An AMOVA of cpDNA variation indicated differentiation between all dimorphic and trimorphic populations. However, this was due to genetic divergence between the two varieties and cpDNA variation was not associated with sexual system within var. *triandrus*. Similarly, AMOVA of microsatellite variation also revealed differentiation between all dimorphic and trimorphic populations but not within populations sampled from the northwestern part of the range. Therefore, although geographically-based population structure is evident in *N. triandrus*, differences in sexual system do not appear to play an important role in genetic differentiation. This pattern suggests that contemporary selective forces maintain differentiation between the sexual systems of *N. triandrus*.

Although there was no evidence of genetic differences between dimorphic and trimorphic populations of *N. triandrus* at microsatellite loci, partial Mantel tests identified a significant association between pairwise differences in M-morph frequency and genetic differentiation. This pattern was evident among all populations in our study and also those sampled from the northwestern portion of the range (Fig. 4). It is unclear what processes are responsible for this pattern and both selective and non-adaptive hypotheses may be involved. For example, there may be selection against immigrants from populations with divergent M-morph frequencies because morph frequencies are associated with morphological variation that is adapted to particular ecological conditions (Barrett *et al*. 2004). The high mutation rates of microsatellites should reveal an effect of historical relatedness among populations on more recent time scales than would be evident from cpDNA markers. Therefore, populations with similar M-morph frequencies could be more closely related to one another. Studies at other nuclear loci may provide further insight into the population structure of *N. triandrus* and the evolutionary origins of dimorphic and trimorphic sexual systems.

Acknowledgements

We thank Adriana Puentes, Erin Ellis, and Pauline Wang for technical help, Bill Cole, Reagan Johnson for field assistance, Marcel Dorken and Melanie Glaettli for their comments on the manuscript and Asher Cutter and Aneil Agrawal for helpful discussions. This work was funded by a Canada Research Chair to SCHB and the Natural Sciences and Engineering Research Council of Canada through a Discovery Grant to SCHB and a postgraduate fellowship to KAH.

References

Baker HG (1959) The contribution of autecological and genecological studies to our knowledge of the past migration of plants. *American Naturalist*, **93**, 255–272.

Barrett SCH (1992) *Evolution and Function of Heterostyly*. Springer-Verlag, Berlin.

Barrett SCH, Forno IW (1982) Style morph distribution in new world populations of *Eichhornia crassipes* (Mart) Solms-Laubach (Water Hyacinth). *Aquatic Botany*, **13**, 299–306.

Barrett SCH, Hodgins KA (2006) Floral design and the evolution of asymmetrical mating systems. In: *Ecology and Evolution of Flowers* (eds Harder LD, Barrett SCH), pp. 239–254. Oxford University Press, Oxford, UK.

Barrett SCH, Cole WW, Arroyo J, Cruzan MB, Lloyd DG (1997) Sexual polymorphisms in *Narcissus triandrus* (Amaryllidaceae): is this species tristylous? *Heredity*, **78**, 135–145.

Barrett SCH, Dorken ME, Case AL (2001) A geographical context for the evolution of plant reproductive systems. In: *Integrating Ecological and Evolutionary Processes in a Spatial Context* (eds Silvertown J, Antonovics J), pp. 341–364. Blackwell, Oxford. UK.

Barrett SCH, Harder LD, Cole WW (2004) Correlated evolution of floral morphology and mating-type frequencies in a sexually polymorphic plant. *Evolution*, **58**, 964–975.

Blanchard JW (1990) *Narcissus: A Guide to Wild Daffodils*. Alpine Garden Society, Surrey, UK.

Brewer S, Cheddadi R, de Beaulieu JL, Reille M (2002) The spread of deciduous *Quercus* throughout Europe since the last glacial period. *Forest Ecology Management*, **156**, 27–48.

Brookfield JFY (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology*, **5**, 453–455.

Charlesworth D (1979) The evolution and breakdown of tristyly. *Evolution*, **33**, 486–498.

Comes HP, Abbott RJ (2000) Random amplified polymorphic DNA (RAPD) and quantitative trait analyses across a major phylogeographical break in the Mediterranean ragwort *Senecio gallicus* Vill. (Asteraceae). *Molecular Ecology*, **9**, 61–76.

Darwin C (1877) *The Different Forms of Flowers on Plants of the Same Species*. John Murray, London, UK.

Demesure B, Sodzi N, Petit RJ (1995) A set of universal primers for amplification of polymorphic noncoding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology*, **4**, 129–131.

Dorken ME, Barrett SCH (2004) Chloroplast haplotype variation among monoecious and dioecious populations of *Sagittaria latifolia* (Alismataceae) in eastern North America. *Molecular Ecology*, **13**, 2699–2707.

Dumolin-Lapegue S, Pemonge MH, Petit RJ (1997) An enlarged set of consensus primers for the study of organelle DNA in plants. *Molecular Ecology*, **6**, 393–397.

Eckert CG, Barrett SCH (1995) Style morph ratios in tristylous *Decodon verticillatus* (Lythraceae): selection vs historical contingency. *Ecology*, **76**, 1051–1066.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

Ennos AR (1994) Estimating the relative rates of pollen and seed migration among plant populations. *Heredity*, **72**, 250–259.

Epperson BK (2002) Spatial-temporal properties of gene genealogies in geographically structured populations. In: *Modern Developments in Theoretical Population Genetics* (eds Slatkin M, Veuille M), pp. 165–182. Oxford University Press, Oxford, UK.

Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, **11**, 1591–1604.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

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

Fernandes A (1965) Contribution à la connaissance de la génétique de l'hétérostylie chez le genre *Narcissus L*. II. l'hétérostylie chez quelques populations de *N. triandrus* var *cernuus* et *N. triandrus* var *concolor*. *Genética Ibérica*, **17**, 215–239.

Garcia-Paris M, Alcobendas M, Buckley D, Wake DB (2003) Dispersal of viviparity across contact zones in Iberian populations of fire salamanders (*Salamandra*) inferred from discordance of genetic and morphological traits. *Evolution*, **57**, 129–143.

Geber MA, Dawson TE, Delph LF (1999) *Gender and Sexual Dimorphism in Flowering Plants*. Springer-Verlag, Berlin.

Gómez A, Lunt DH (2006) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: *Phylogeography in Southern European Refugia: Evolutionary Perspectives on the Origins and Conservation of European Biodiversity* (eds Weiss S, Ferrand N), pp. 155–182. Springer, Amsterdam, the Netherlands.

Goodman SJ (1997) R_{ST} CALC: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Molecular Ecology*, **6**, 881–885.

Goudet J (2001) FSTAT: *A Program to Estimate and Test Gene Diversities and Fixation Indices*, Version 2.9.3. www2.unil.ch/popgen/softwares/fstat.htm.

Gubitz T, Thorpe RS, Malhotra A (2000) Phylogeography and natural selection in the Tenerife gecko *Tarentola delalandii*: testing historical and adaptive hypotheses. *Molecular Ecology*, **9**, 1213–1221.

Hamrick JL, Godt MJ (1989) Allozyme diversity in plant species. In: *Plant Population Genetics, Breeding, and Genetic Resources* (eds Brown AHD, Clegg MT, Kahler AL, Weir BS), pp. 43–63. Sinauer, Sunderland, Massachusetts.

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

Hardy OJ, Vekemans X (2002) SPAGEDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618–620.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hodgins KA, Barrett SCH (2006a) Female reproductive success in the evolution of mating-type frequencies in tristylous populations. *New Phytologist*, **171**, 569–580.

Hodgins KA, Barrett SCH (2006b) Mating patterns and demography and the tristylous daffodil *Narcissus triandrus*. *Heredity*, **96**, 262–270.

Hodgins KA, Stehlik I, Wang P, Barrett SCH (2006) The development of eight microsatellite loci in the wild daffodil *Narcissus triandrus* (Amaryllidaceae). *Molecular Ecology Notes*, doi: 10.1111/j.1471-8286.2006.01637.x

Huntley B, Birks JDS (1983) *An Atlas of Past and Present Pollen Maps for Europe: 0–13 000 Years Ago*. Cambridge University Press, Cambridge.

Knowles LL, Maddison WP (2002) Statistical phylogeography. *Molecular Ecology*, **11**, 2623–2635.

Martinez-Solano I, Teixeira J, Buckley D, Garcia-Paris M (2006) Mitochondrial DNA phylogeography of *Lissotriton boscai* (Caudata, Salamandridae): evidence for old, multiple refugia in an Iberian endemic. *Molecular Ecology*, **15**, 3375–3388.

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular-data.2. Gene-frequency data. *Journal of Molecular Evolution*, **19**, 153–170.

Olalde M, Herran A, Espinel S, Goicoechea PG (2002) White oaks phylogeography in the Iberian Peninsula. *Forest Ecology and Management*, **156**, 89–102.

Ota T (1993) DISPAN: *Genetic Distance and Phylogenetic Analysis*, Version 1.1. Pennsylvania State University, University Park.

Paulo OS, Jordan WC, Bruford MW, Nichols RA (2002) Using nested clade analysis to assess the history of colonisation and the persistence of populations of an Iberian lizard. *Molecular Ecology*, **11**, 809–819.

Pérez R, Vargas P, Arroyo J (2004) Convergent evolution of flower polymorphism in *Narcissus* (Amaryllidaceae). *New Phytologist*, **161**, 235–252.

Petit RJ, Grivet D (2002) Optimal randomization strategies when testing the existence of a phylogeographic structure. *Genetics*, **161**, 469–471.

Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, **12**, 844–855.

Petit RJ, Aguinagalde I, de Beaulieu JL *et al*. (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, **300**, 1563–1565.

Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin GG (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, **14**, 689–701.

Posada D, Crandall KA, Templeton AR (2000) GEODIS: a program for the cladistic nested analysis of geographic distribution of genetic haplotypes. *Molecular Ecology*, **9**, 487–488.

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

Rieseberg LH, Beckstromsternberg S, Doan K (1990) *Helianthus annuus* ssp *texanus* has chloroplast DNA and nuclear ribosomal RNA genes of *Helianthus debilis* ssp *cucumerifolius*. *Proceedings of the National Academy of Sciences of the USA*, **87**, 593–597.

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

Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN: *A Software for Population Genetics Data Analysis*, Version 2.000. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Geneva, Swtizerland.

Sculthorpe CD (1967) *The Biology of Aquatic Vascular Plants*. Edward Arnold, London, UK.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology*, **17**, 1105–1109.

Templeton AR (1998) Nested clade analysis of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology*, **7**, 381–397.

Templeton AR (2002) 'Optimal' randomization strategies when testing the existence of a phylogeographic structure: a reply to Petit and Grivet. *Genetics*, **161**, 473–475.

Templeton AR (2004) Statistical phylogeography: methods of evaluating and minimizing inference errors. *Molecular Ecology*, **13**, 789–809.

Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, **132**, 619–633.

Templeton AR, Routman E, Phillips CA (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics*, **140**, 767–782.

Terauchi R, Terachi T, Tsunewaki K (1991) Intraspecific variation of chloroplast DNA in *Dioscorea bulbifera* L. *Theoretical and Applied Genetics*, **81**, 461–470.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

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

Zhang DX, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology*, **12**, 563–584.

This work was part of Kathryn Hodgins' Ph.D. thesis research at the University of Toronto investigating the evolution and maintenance of sexual systems in the wild daffodil *Narcissus triandrus*. Hodgins' main interests are in evolutionary genetics and particularly understanding mechanisms responsible for evolutionary transitions between sexual systems. Spencer Barrett works on the evolution and ecology of reproductive systems and the genetics of colonization and migration.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd