

## Genetic diversity in *Cypripedium calceolus* (Orchidaceae) with a focus on north-western Europe, as revealed by plastid DNA length polymorphisms

Michael F. Fay<sup>1,\*</sup>, Ruth Bone<sup>1</sup>, Peter Cook<sup>1</sup>, Imalka Kahandawala<sup>1,2</sup>, Jennifer Greensmith<sup>1</sup>, Stacey Harris<sup>1</sup>, Henrik Æ. Pedersen<sup>3</sup>, Martin J. Ingrouille<sup>2</sup> and Christian Lexer<sup>1,4</sup>

<sup>1</sup>Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, UK, <sup>2</sup>Birkbeck, University of London, Malet Street, London WC1E 7HX, UK, <sup>3</sup>Botanical Garden & Museum, Natural History Museum of Denmark, University of Copenhagen, Gothersgade 130, DK-1123 Copenhagen K, Denmark and <sup>4</sup>University of Fribourg, Department of Biology, Unit of Ecology & Evolution, Chemin du Musée 10, CH-1700 Fribourg, Switzerland

Received: 16 October 2008 Returned for revision: 28 November 2008 Accepted: 7 April 2009 Published electronically: 19 May 2009

• **Background and Aims** *Cypripedium calceolus*, although widespread in Eurasia, is rare in many countries in which it occurs. Population genetics studies with nuclear DNA markers on this species have been hampered by its large nuclear genome size. Plastid DNA markers are used here to gain an understanding of variation within and between populations and of biogeographical patterns.

• **Methods** Thirteen length-variable regions (microsatellites and insertions/deletions) were identified in non-coding plastid DNA. These and a previously identified complex microsatellite in the *trnL-trnF* intergenic spacer were used to identify plastid DNA haplotypes for European samples, with sampling focused on England, Denmark and Sweden.

• **Key Results** The 13 additional length-variable regions identified were two homopolymer (polyA) repeats in the *rps16* intron and a homopolymer (polyA) repeat and ten indels in the *accD-psaI* intergenic spacer. In *accD-psaI*, most of these were in an extremely AT-rich region, and it was not possible to design primers in the flanking regions; therefore, the whole intergenic spacer was sequenced. Together, these new regions and the *trnL-trnF* complex microsatellite allowed 23 haplotypes to be characterized. Many were found in only one or a few samples (probably due to low sampling density), but some commoner haplotypes were widespread. Most of the genetic variation was found within rather than between populations (83 vs. 18%, respectively). Two haplotypes occurred from the Spanish Pyrenees to Sweden.

• **Conclusions** Plastid DNA data can be used to gain an understanding of patterns of genetic variation and seed-mediated gene flow in orchids. Although these data are less information-rich than those for nuclear DNA, they present a useful option for studying species with large genomes. Here they support the hypothesis of long-distance seed dispersal often proposed for orchids.

**Key words:** Biogeography, *Cypripedium calceolus*, genome size, plastid microsatellites, population genetics, seed dispersal.

### INTRODUCTION

*Cypripedium calceolus* is one of the rarest plant species in the UK, with only one surviving original clump and some reintroduced plants present in the wild in northern England. It was formerly more widespread, although never common, and the reduction in its populations to near extinction has been attributed to the collection of plants for gardens and herbarium specimens (Nelson, 1994; Farrell, 1999). In addition to the plants in the wild, there are several plants of known UK origin in cultivation and three plants growing in semi-natural habitat which have been considered to be possible introductions from non-native material (Farrell, 1999; I. Taylor, Natural England, pers. comm.). One of these has been shown to be a member of the North American *C. parviflorum* complex (M. F. Fay *et al.*, unpubl. res.), and is excluded from the present study.

Due to its extreme rarity in England, *C. calceolus* has been considered an important target for conservation activities

ranging from site management, wardening, pollination, *ex situ* seed germination and, more recently, genetic fingerprinting. However, largely due to the large genome size of this species (1C = 32.4 pg, Bennett *et al.*, 2000), the nuclear DNA-based techniques increasingly used for studies of rare and endangered species (Qamaruz-Zaman *et al.*, 1998; Fay and Krauss, 2003), including amplified fragment length polymorphisms (AFLP, Vos *et al.*, 1995) and random amplified polymorphic DNA (RAPD, Williams *et al.*, 1990), have not proved effective (Fay and Cowan, 2001; Fay *et al.*, 2002; Fay and Krauss, 2003). When the standard AFLP protocol is used with genomes as large as that in *C. calceolus*, levels of genetic variation are dramatically underestimated (Fay and Cowan, 2001; Fay *et al.*, 2002, 2005; Leitch and Fay, 2008). Nuclear microsatellites, similarly, have been shown to be difficult to apply to organisms with large genomes (Garner, 2002).

To overcome this problem, Fay and Cowan (2001) investigated the use of plastid microsatellites as alternative markers. The plastid genome comprises a single, circular

\* For correspondence. E-mail [m.fay@kew.org](mailto:m.fay@kew.org)

chromosome. Unlike the nuclear chromosomes, this exists as a single type inherited from one parent only (the female in orchids and most other angiosperms; Corriveau and Coleman, 1988) and shows relatively low levels of size variation (Chase and Palmer, 1989). The sequence of genes is well characterized, and in many cases the genes contain introns and/or are separated by intergenic spacers (IGS), both non-coding regions of DNA. At least some of these non-coding regions contain repetitive motifs, commonly consisting of either all adenine (A) or all thymine (T) residues or a combination of these two bases. When these regions contain higher numbers of repeating units (generally ten or more for single nucleotide repeats), they are referred to as plastid microsatellites. They show high levels of variability, and the length of a microsatellite can vary extensively within and between populations. They have proved especially useful in studies of biogeography, and in Europe they have been used to study biogeographical patterns in a range of orchids, including *Orchis* spp. (Qamaruz-Zaman, 2000; Fay et al., 2007; Bateman et al., 2008), *Dactylorhiza* spp. (Shipunov et al., 2004, 2005; Pillon et al., 2007) and *Spiranthes romanzoffiana* (Forrest et al., 2004). In addition, non-coding plastid DNA often also includes insertions/deletions (indels) due to tandem repeats of longer motifs (minisatellites) that can vary in length from a few to hundreds of base pairs (bp), for example in *Sorbus* (King and Ferris, 2002) and *Anacamptis* (Cozzolino et al., 2004).

The placement of these length-variable regions varies from taxon to taxon, and an initial survey of plastid DNA is necessary to locate them. Once suitable regions are identified, primers are designed to amplify fragments containing them. Amplification products are obtained for all individuals under study, and their length in base pairs is determined. This technique has a general advantage over multilocus genetic fingerprinting methods such as RAPD and AFLP in that only one short region is amplified per reaction, making the technique less sensitive to DNA quality and quantity. It is therefore, at least potentially, applicable to DNA samples extracted under less than ideal circumstances, e.g. from herbarium material (Fay and Cowan, 2001; Cozzolino et al., 2007).

Here we report the development of a set of length-variable fragments of plastid DNA for *C. calceolus* and their use in analysing patterns of genetic variation in this species. For plants in England, these data are used to infer the origin of putatively non-native material and to inform the choice of material to be used for reintroduction purposes.

## MATERIALS AND METHODS

### *Plant materials*

The plant material used in this study is listed in Table 1. Detailed localities are not given, due to the sensitivity surrounding many of the sites. New specimens were collected into silica gel using the method of Chase and Hills (1991). These included the one remaining wild plant and all living plants of known wild origin in England. Representative samples from elsewhere in Europe were included for comparative purposes. In Denmark, only two populations are known, and 34 plants were sampled from these. Because these had

previously been shown to be genetically invariant (I. Kahandawala et al., unpubl. res.), however, only three representative samples are included from each population here. DNA was extracted from silica-gel dried leaf material, using a modified 2× CTAB (cetyltrimethyl-ammonium bromide) procedure (Doyle and Doyle, 1987) followed by purification on a caesium chloride gradient or columns following the manufacturers' protocols.

### *Identification of length-variable regions*

Length-variable regions flanked by conserved regions were sought in the *accD-psaI* IGS and the *rps16* intron. These were amplified using primers (Table 2) given in Small et al. (1998) and Oxelman et al. (1997), respectively, from three representative samples (English, Polish and Swiss) of *C. calceolus* and sequenced using modified dideoxy cycle sequencing with dye terminators run on an ABI 377 automated sequencer. The sequences were aligned manually in PAUP\* 4.0b4 (Swofford, 2002), and variable regions were identified. Primers to amplify the length-variable regions in the *rps16* intron were designed in conserved flanking regions (Table 3). For each region, one primer was labelled with a fluorescent dye to enable the amplification products to be visualized using an ABI 377 automated sequencer. Sizes (bp) were determined using GeneScan 3.1 and Genotyper 2.0 (Applied Biosystems Inc., Foster City, CA, USA). Due to the AT-rich nature of the flanking regions for the length-variable regions in *accD-psaI* (see Results), primers could not be designed to amplify these regions. The entire IGS was therefore sequenced for all individuals to allow these regions to be scored. Sequences containing unique or rare indels were redone from

TABLE 1. *Plant material used in this study*

Origin*	Sample size	Notes
Austria, Steiermark (C)	5	
Denmark, Jutland (N)	6	Two localities (three individuals from each)
Estonia, Muhu (C)	1	
France, SE (W)	3	
Italy, Dolomites (C)	1	
Poland, West (C)	4	
Poland, North East (C)	2	
Spain, Pyrenees (W)	9	
Sweden, Dalarna (N)	37	10 localities
Sweden, Jämtland (N)	3	
Switzerland, Canton Bern (C)	6	
UK, England (W)	7	Wild site (two shoots were both sampled) <sup>†</sup> plus five plants in semi-natural sites or cultivation

Sample size refers to the number of individuals for which haplotype data were collected. Due to small sample sizes, for statistical analyses Austria and Italy were combined as 'Eastern Alps' and Estonia and Poland were combined as 'Poland/Estonia'.

\* C, N and W stand for Central, Northern and Western Europe, as used in the text.

<sup>†</sup> At the wild site, the remaining plant had two main shoots. Both of these were sampled to investigate the possibility that there might be two plants. No genetic differences were detected, and we assume that it is actually one genet.

TABLE 2. Primers used to amplify the *accD-psaI* intergenic spacer and the *rps16* intron

Locus	Primer	Sequence (5' to 3')	Reference
<i>accD-psaI</i> intergenic spacer	<i>accD</i> -769F	GGAAGTTTGAGCTTATGCAAATGG	Small <i>et al.</i> (1998)
<i>rps16</i> intron	<i>psaI</i> -75R	AGAAGCCATTGCAATTGCCGGAAA	Oxelman <i>et al.</i> (1997)
	<i>rps16</i> -1F	CACGGGTCGCCCTCGTTCCG	
	<i>rps16</i> -1R	TCGGGATCGAACATCAATTGCAAC	

TABLE 3. Primer sequences used to amplify the two (A)<sub>n</sub> microsatellites in the *rps16* intron and the complex microsatellite in the *trnL-trnF* intergenic spacer (*cyp2*)

Region	Primer	Sequence (5' to 3')	Reference
<i>rps16</i> -1	<i>rps16</i> -1F	CTTGCTCAACAATTGGATC	This paper
	<i>rps16</i> -1R	CGTACTCCTAACTCAAGTTG	This paper
<i>rps16</i> -2	<i>rps16</i> -2F	CAACTTGAGTTAGGAGTACG	This paper
	<i>rps16</i> -2R*	See Table 2	Oxelman <i>et al.</i> (1997)
<i>cyp2</i>	<i>trnL-trnF</i>	GGTTCAAGTCCCTCTATCCC	Taberlet <i>et al.</i> (1991)
	<i>cyp2</i> -R	CGGATCCATTTGTAAAGAACAG	Fay and Cowan (2001)

\* The original *rps16* reverse primer for amplification of the whole locus was used due to the proximity of the microsatellite region to the end of the amplified sequence.

new PCR products to test for reproducibility and to exclude possible *Taq*-induced artefacts. The new regions identified here and one identified by Fay and Cowan (2001) in the *trnL-trnF* IGS (*cyp2*) were scored for all individuals. [The other region in the *trnL* intron, *orch1*, used by Fay and Cowan (2001), proved to be invariant in a subset of the samples and was not used further.]

A matrix was prepared, using different numbers for alleles of different lengths of the microsatellite regions and 1 vs. 2 (see Table 4) for the simple indels. Where two or more indel events occurred at the same position in different samples (presumed to be single indels of different lengths rather than sequential indels), the alleles were numbered from 1 to 3 or 4, as appropriate (see Table 4).

#### Definition of haplotypes

Haplotypes numbered 1–23 were defined by the combination of the alleles for the 14 different loci. The haplotype definitions are given in Table 5.

#### Statistical data analysis

Haplotype frequencies in the sampled populations were estimated using ARLEQUIN software (Schneider *et al.*, 2000). Genetic diversity was estimated as the number of haplotypes, number of polymorphic sites and gene diversity using ARLEQUIN and haplotype richness following Petit *et al.* (1998) using the CONTRIB program. Haplotype richness was corrected for differences in sample size using the rarefaction method. To provide robust estimates, the sample size of the smallest population sample ( $n = 3$ ) was used for rarefaction. The distribution of plastid DNA diversity within and among populations was studied using analysis of molecular variance (AMOVA, Excoffier *et al.*, 1992) as implemented in

TABLE 4. Description of the 14 loci, the position of the indels in the aligned *accD-psaI* matrix, the alleles found and a description of alleles for the insertions/deletions

Locus	Type	Position <sup>†</sup>	Alleles	Allele description <sup>‡</sup>
1	<i>accD-psaI</i> microsatellite (A) <sub>n</sub>	–	1,2,3,4	
2	<i>accD-psaI</i> indel 1	350	1,2,3	long insert–no insert–short insert
3	<i>accD-psaI</i> indel 2	400	1,2,3,4	long–middle–short–none
4	<i>accD-psaI</i> indel 3	465	1,2	normal–deletion
5	<i>accD-psaI</i> indel 4	522	1,2,3	insert–normal–deletion
6	<i>accD-psaI</i> indel 5	630	1,2	normal–insert
7	<i>accD-psaI</i> indel 6	750	1,2,3	normal–short insert–long insert
8	<i>accD-psaI</i> indel 7	812	1,2	normal–insert
9	<i>accD-psaI</i> indel 8	895	1,2	normal–deletion
10	<i>accD-psaI</i> indel 9	1050	1,2	normal–deletion
11	<i>accD-psaI</i> indel 10	1208	1,2	normal–insert
12	<i>rps16</i> -1 microsatellite (A) <sub>n</sub>	–	1,2	
13	<i>rps16</i> -2 microsatellite (A) <sub>n</sub>	–	1,2,3	
14	<i>cyp2</i> microsatellite (AT) <sub>n</sub> T <sub>n</sub>	–	1,2,3,4	

<sup>†</sup> Position in base pairs in aligned *accD-psaI* matrix.

<sup>‡</sup> ‘Normal’ refers to the condition found in the majority of samples.

Directionality of insertions and deletions is inferred, with the commonest allele being treated as ‘normal’.

ARLEQUIN, and the significance of the AMOVA was tested based on 10 000 permutations of haplotypes among populations. A comparison of the distribution of diversity for ordered vs. unordered haplotypes was carried out following Pons and Petit (1996) using the PERMUT/cpSSR software. Two different methods were used for obtaining ordered

TABLE 5. *The allelic composition of haplotypes 1–23 (the alleles are described in Table 4)*

Haplotype	Alleles
H1	12112121111123
H2	22112111112113
H3	31112121111123
H4	32112121222113
H5	32112121112114
H6	32112121111122
H7	32112121111123
H8	32112121222213
H9	32112131111122
H10	32113121111122
H11	32212121111133
H12	32322121111113
H13	33111121111123
H14	33112121111123
H15	32111121111123
H16	42112121111123
H17	42111121111124
H18	42112121111124
H19	42322122111113
H20	42322122111114
H21	42322122121114
H22	42412221111133
H23	43112121111123

haplotype information: (1) using the number of mutation steps between haplotypes (genetic divergence estimated as  $N_{ST}$ ), and (2) making use of information on repeat number of each polymorphic site (genetic divergence estimated as  $R_{ST}$ ). The latter method was specifically designed for microsatellites, but the former method makes fewer assumptions about the nature and complexity of mutational patterns detected by fragment length analysis (Pons and Petit, 1996). A median-joining (MJ) network (Bandelt *et al.*, 1999) was constructed based on plastid DNA haplotypes using the program NETWORK 4.2.0.1. ([www.fluxus-engineering.com](http://www.fluxus-engineering.com)). This method is able to resolve even complex haplotype patterns (Posada and Crandall, 2001). It uses parsimony criteria to identify median vectors, i.e. consensus sequences of mutually close sequences of markers that are biologically equivalent to possible unsampled or extinct ancestral haplotypes. To reduce complexity of the network, rapidly mutating characters were downweighed in the analysis.

## RESULTS

### *Identification of length-variable regions*

Two homopolymer microsatellites (both A on the sense strand) were detected in the *rps16* intron, each consisting of 11 or 12 bp in the samples studied. In the full sampling, ten indels ranging in size from 6 to 21 bp and one homopolymer microsatellite (also A on the sense strand) of 11–12 bp were found in the *accD-psaI* IGS. The amplified region of the *accD-psaI* IGS was exceedingly AT-rich (approx. 80 %, depending on the sample) and most of the length-variable regions were found in a fragment of 461 bp in the aligned matrix (positions 328–788), in which only one base (a G at position 581) was

not A or T. In all cases where sequences were redone from new PCR products, all the same indels were found. Representative sequences are available from the first author ([m.fay@kew.org](mailto:m.fay@kew.org)).

### *Haplotypes recovered*

All 14 length-variable regions were successfully scored for all samples, and all loci were variable. The final matrix contained information for 84 (111 if all 34 Danish samples were included) individuals. The alleles found are presented in Table 4. Twenty-three haplotypes were obtained from the combined alleles for the 14 regions (Table 5).

### *Distribution of haplotypes*

Haplotype H6 was the most common. Haplotypes H14 and H16 were also relatively common, and these three haplotypes were also widespread (see Fig. 1, Table 6). Two (H6 and H16) were found in individuals ranging from the Spanish Pyrenees to northern Sweden. Other haplotypes, at this level of sampling, were more localized, with some only being represented by one individual.

### *Distribution of genetic diversity in populations*

The more diverse populations in terms of gene diversity and haplotype richness were Eastern Alps, France, Poland/Estonia and Switzerland (Table 6). If haplotype numbers and numbers of polymorphic sites are considered in addition, then Sweden/Dalarna also emerges as being among the more variable populations (eight haplotypes found among 37 specimens, eight polymorphic sites; Table 7). A high proportion of the genetic variability in the haplotype data was attributed to the within-population level (82.5 %) whereas only 17.5 % resided among populations (Table 8). Genetic divergence estimated with ordered haplotypes was not significantly different from divergence for unordered haplotypes, regardless of whether  $R_{ST}$  or  $N_{ST}$  was used to estimate divergence for ordered haplotypes. The results for the  $N_{ST}$  analysis are shown in Table 9.

### *Haplotype network*

The MJ network was composed of a predominantly Western European torso of closely related haplotypes (just one mutation step between nodes) and a number of predominantly Central European branches made up of more divergent haplotypes (up to four mutation steps between nodes; Fig. 1). Haplotypes H6, H14 and H16 were the most common. Each of these haplotypes was found in a central position in the network (at least three connections branching out from each haplotype), thus indicating that these are older, ancestral haplotypes that gave rise to less common, derived haplotypes found elsewhere in the network.

### *Haplotypes found in English individuals*

Five haplotypes (H2, H10, H14, H17 and H23) were found among individuals from England (Table 6). Of these, H14 was widespread, also being found in France, Switzerland and





TABLE 6. *Haplotype frequencies in nine populations of Cypripedium calceolus, determined by sequence or fragment analysis of 14 polymorphic plastid DNA sites*

Haplotype	Eastern Alps	Denmark	France	Poland/Estonia	Spain	Sweden/Dalarna	Sweden/Jämtland	Switzerland	UK
H1	0	0	0	0	0	1	0	0	0
H2 <sup>†</sup>	0	0	0	0	0	0	0	0	1
H3	1	0	0	0	0	0	0	0	0
H4	0	0	0	1	0	5	0	0	0
H5	0	0	0	1	0	0	0	0	0
H6	0	0	0	3	5	14	0	2	0
H7	0	0	0	0	0	5	0	0	0
H8	0	0	0	0	0	1	0	0	0
H9	0	0	0	0	0	0	0	1	0
H10	0	0	1	0	1	0	0	1	1
H11	1	0	0	0	0	0	0	0	0
H12	0	0	0	1	0	0	0	0	0
H13	0	0	1	0	0	0	0	0	0
H14	0	0	1	0	0	8	3	1	3
H15	0	0	0	1	0	0	0	0	0
H16	0	6	0	0	2	2	0	0	0
H17	0	0	0	0	1	0	0	0	1
H18	1	0	0	0	0	0	0	0	0
H19	1	0	0	0	0	0	0	0	0
H20	0	0	0	0	0	0	0	1	0
H21	1	0	0	0	0	0	0	0	0
H22	1	0	0	0	0	0	0	0	0
H23	0	0	0	0	0	1	0	0	1

\* For more information on populations sampled see Table 1; for more information on polymorphic plastid DNA sites analysed see Table 4.

<sup>†</sup> Haplotype H2, found exclusively in Britain, has previously been suspected to be of introduced origin. This haplotype was included in the computation of the median-joining (MJ) network shown in Fig. 1, but not in analyses of plastid DNA diversity and divergence shown in Tables 7–9.

TABLE 7. *Plastid DNA diversity in populations of Cypripedium calceolus, including for each population the number of haplotypes found (sample size), number of polymorphic sites, genetic diversity with standard error and allelic richness*

Population	No. of haplotypes (sample size)	No. of polymorphic sites	Gene diversity ( $\pm$ s.e.)	Allelic richness
Eastern Alps	6 (6)	9	1.000 (0.096)	2.0
Denmark	1 (6)	0	0.000 (0.000)	0.0
France	3 (3)	3	1.000 (0.272)	2.0
Poland/Estonia	5 (7)	9	0.857 (0.137)	1.6
Spain	4 (9)	3	0.694 (0.147)	1.2
Sweden/Dalarna	8 (37)	8	0.790 (0.044)	1.4
Sweden/Jämtland	1 (3)	0	0.000 (0.000)	0.0
Switzerland	5 (6)	9	0.933 (0.122)	1.8
Britain	4 (6)	4	0.800 (0.172)	1.5

\* Corrected for differences in sample size by rarefaction for  $n = 3$ , the smallest sample size among the populations studied.

## DISCUSSION

### *Utility of plastid DNA polymorphisms in Cypripedium calceolus*

Plastid DNA polymorphisms are here shown to have the capacity to reveal genetic variability in *C. calceolus*, providing an alternative to the use of nuclear DNA, which in this species is compromised by the large nuclear genome size. Although data from plastid DNA are generally less information-rich due to the uniparental pattern of inheritance, in this case genetic fingerprinting using AFLP appears to have given a dramatic underestimate of levels of genetic variation (Fay and

Cowan, 2001; Fay *et al.*, 2002, 2005; Leitch and Fay, 2008). In a separate study, aiming at developing nuclear microsatellites for *C. calceolus*, only two out of seven which gave scorable results were polymorphic (I. Kahandawala *et al.*, unpubl. res.), possibly indicating a similar problem.

The A/T-rich region of the *accD-psaI* IGS was particularly informative in this case, and the extra effort required in terms of sequencing the fragment rather than sizing it (as for the other loci) was repaid by the increase in the number of haplotypes revealed. The percentage A/T approaches 100% for 461 bp, making this the most A/T-rich region of DNA of which we are aware.

### *Distribution of genetic variation*

Central European material, despite relatively sparse sampling, appears to be genetically more variable than northern and western material. In principle, the high levels of genetic diversity in the Eastern Alps and Switzerland (Table 7) are compatible with two hypotheses: (1) the presence of refugial populations in Central Europe, in regions surrounding the ice sheets that covered the Alps during the Weichselian glaciation; and (2) genetic admixture in this region involving two or more postglacial colonization routes from unsampled southern/south-eastern European refugia. A study of genetic structure with nuclear markers, including the estimation of admixture proportions for individual plants, would provide additional insight (although this may be difficult to achieve due to the problems with genome size discussed above).

The high level of genetic diversity detected in Poland/Estonia (Table 7) corroborates the findings of Kull & Paaver

TABLE 8. *Distribution of plastid DNA variation within and among populations of Cypripedium calceolus, estimated by Analysis of Molecular Variance (AMOVA)*

Source of variation	Sums of squares	Variance components	d.f.	Variation (%)	F-statistic	P*
Among populations	24.5	0.244	8	17.5	0.175	<0.001
Within populations	84.7	1.145	74	82.5		

\* The significance threshold was determined by 10 000 permutations of haplotypes among populations.

TABLE 9. *Analysis of plastid DNA diversity following Pons & Petit (1996), including the number of populations studied, number of haplotypes considered, and diversity and divergence statistics for unordered and ordered haplotypes*

No. of populations	No. of haplotypes	$h_S^*$	$h_T^*$	$G_{ST}^*$	$v_S^\dagger$	$v_T^\dagger$	$N_{ST}^\dagger$
9	22	0.675	0.907	0.256	0.694	0.905	0.233

\* Unordered haplotypes:  $h_S$ , within-population diversity;  $h_T$ , total diversity;  $G_{ST}$ , genetic divergence.

† Ordered haplotypes:  $v_S$ , within-population diversity;  $v_T$ , total diversity;  $N_{ST}$ , genetic divergence. The difference in genetic divergence when estimated using unordered vs. ordered haplotype information ( $G_{ST}$  vs.  $N_{ST}$ ) was not significant.

(1997) and Brzosko *et al.* (2002). The former, studying three strongly polymorphic aspartate aminotransferase loci, observed high levels of heterozygosity in seven Estonian populations of *C. calceolus* ( $H_o = 0.40-0.53$ ). An eighth population was less genetically diverse ( $H_o = 0.16$ ), which was ascribed to a founder effect. Brzosko *et al.* (2002) examined 11 allozyme loci in three populations from the Biebrza valley, north-eastern Poland, and assessed the genetic diversity as percentage polymorphic loci ( $P$ ), mean number of alleles per locus ( $A$ ), and observed ( $H_o$ ) and expected ( $H_e$ ) heterozygosity. They found diversity within populations to be relatively high compared with rare taxa and taxa with similar life histories ( $P = 45.5\%$  in all populations;  $A = 0.184$ ;  $H_o = 0.184$  on average;  $H_e = 0.156$  on average). The genetic variation found between populations was even smaller (1.6%) than in the present study.

The long branches in the predominantly Central European section of the haplotype network (Fig. 1) may be indicative of the presence of genetic material derived from a different post-glacial colonization route (or routes), possibly stemming from unsampled populations to the east of the region on which the present study was focused. Central European and eastern material remains undersampled and is an obvious focus for future studies, although the difficulty of obtaining samples due to legal protection of this species in many countries may make this problematic. This would potentially clarify the number of refugia that existed for this species during the last ice age. The variability found in Swedish material may indicate that Sweden was recolonized by individuals from two or more refugia, migrating along western and central/eastern routes.

The absence of a significant difference between the divergences for ordered haplotypes and unordered haplotypes (Table 9) indicates the absence of marked phylogeographical structure in the data, i.e. seed-mediated gene flow was more important than mutation in generating the observed genetic structure. This may obscure the geographical patterns, even if more central and eastern samples are obtained.

#### *Long-distance dispersal of orchids*

Orchid seeds are commonly said to be capable of long-distance dispersal by the wind due to their small size. Darwin (1877, pp. 278–279) stated that ‘The minute seeds within their light coats are well fitted for wide dissemination; and I have several times observed seedlings springing up in my orchard and in a newly-planted wood, which must have come from a considerable distance. This was especially the case with *Epipactis latifolia*; and an instance has been recorded by a good observer [Mr. Bree, in ‘Loudon’s Mag. of Nat. Hist.’ vol. ii. 1829, p. 70.] of seedlings of this plant appearing at the distance of between eight and ten miles from any place where it grew.’ Summerhayes (1951) stated that ‘the seeds are certainly very light, and easily carried for long distances by the wind’ and ‘the seeds are shaken out a few at a time, and are distributed far and wide. These seeds, as in the case with others distributed by the wind, are probably carried upwards, by ascending columns of hot air, to very considerable heights, and then transported long distances in the upper atmosphere before sinking slowly to earth again.’ In their extensive review on orchid seeds, Arditti and Ghani (2000) tabulated published dispersal distances of up to 2000 km for orchid seeds. In contrast, some authors have reported that most orchid seeds fall within only a short distance of the source plant. In *Spiranthes spiralis*, the vast majority of seeds were demonstrated to fall within 15 cm of the source plant (Machon *et al.*, 2003). In *Anacamptis morio*, *Dactylorhiza majalis* and *Pseudorchis albida*, Jersáková and Malinová (2007) demonstrated that the proportion of seeds landing more than 1 m away from the source plant approached zero. In *Orchis purpurea*, Jacquemyn *et al.* (2007) demonstrated ‘rather limited seed dispersal distances’ (approx. 4–5 m), and in *Cypripedium macranthos*, Chung *et al.* (2009) stated that ‘most seeds fall close to the maternal plants’ on the basis of the significant fine-scale spatial genetic structure they found. Although orchid seeds do have the potential for dispersal over great distances it is thus

probable that the pattern of dispersal follows a leptokurtic distribution in most (if not all) cases.

In the present study, some of the haplotypes found in *C. calceolus* were widely distributed, in extreme cases from the Pyrenees to northern Sweden. This and the highly homogeneous genetic structure revealed by AMOVA (Table 8) can be interpreted as support for wide-ranging seed dispersal in these light-seeded, wind-dispersed orchids, at least as an occasional (or rare) occurrence. Our data do not allow us to say whether dispersal occurred from the Pyrenees to Sweden in one or more than one step, but the hypervariable nature of the loci involved leads us to believe that it is unlikely to represent more than a few seed dispersal events. An alternative explanation is that both the Pyrenees and Sweden were colonized by individuals from geographically intermediate populations in glacial refugia. This explanation, however, still involves dispersal over substantial distances.

#### *Implications for conservation and reintroduction*

On the basis of the results presented here, English Nature (now Natural England) reached the decision that seedlings resulting from self-pollination of the two putatively introduced individuals of *C. calceolus* or cross pollination of these individuals with other plants in England should be excluded from the reintroduction programme for this species. Despite the extreme demographic bottleneck suffered by this species in England, a higher level of genetic diversity has been maintained than in the larger populations in Denmark. Data for nuclear microsatellites also show that the native English plants are genetically variable, whereas Danish plants are genetically homogeneous (I. Kahandawala *et al.*, unpubl. res.). This should be taken into account for seed storage and other conservation activities. Due to the low level of sampling reported here, it is not possible to make firm recommendations about conservation activities elsewhere in Europe, but it appears that Central (and some Northern) European populations may be more diverse than those in Western Europe. Swedish populations are more diverse than those in Denmark, and this could be due to there being more than one migration route following the last glaciation. At the same time, the lack of genetic polymorphism in Danish populations could be due to a founder effect, as the present occurrence in the region of Jutland was probably founded by long-distance seed dispersal. Thus, the sampled populations (discovered as late as 1884 and 1968, respectively) are the only Danish populations known since 1767 and the only ones ever known from Jutland.

#### ACKNOWLEDGMENTS

This work was funded by the Royal Botanic Gardens, Kew, and English Nature (now part of Natural England) as part of the Biodiversity Action Plan for *C. calceolus*. We thank all colleagues who provided plant material, particularly Fred Campbell (Sweden), Marcin Zych and Maciej Romański (Poland), José M. Iriondo (Spain), Peter Schönschwetter and Andreas Tribsch (Austria) and Tiit Kull (Estonia). We also thank Mark Chase, Phil Cribb and Ilia Leitch (Kew) and Ian Taylor (Natural England) for useful comments on this

project and Olivier Maurin and Robyn Cowan for support in the laboratory.

#### LITERATURE CITED

- Arditti J, Ghani AKA. 2000. Tansley Review No. 110. Numerical and physical properties of orchid seeds and their biological implications. *New Phytologist* **145**: 367–421.
- Bandelt H-J, Forster P, Roehl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**: 37–48.
- Bateman RM, Smith RJ, Fay MF. 2008. Morphometric and population genetic analyses elucidate the origin, evolutionary significance and conservation implications of *Orchis* × *angusticurvis* (*O. purpurea* × *O. simia*), a hybrid orchid new to Britain. *Botanical Journal of the Linnean Society* **157**: 687–711.
- Bennett MD, Bhandol P, Leitch IJ. 2000. Nuclear DNA amounts in angiosperms and their modern uses – 807 new estimates. *Annals of Botany* **86**: 859–909.
- Brzosko E, Ratkiewicz M, Wróblewska A. 2002. Allozyme differentiation and genetic structure of the Lady's slipper (*Cypripedium calceolus*) island populations in north-east Poland. *Botanical Journal of the Linnean Society* **138**: 433–440.
- Chase MW, Hills HG. 1991. Silica gel: an ideal desiccant for preserving field-collected leaves for use in molecular studies. *Taxon* **41**: 215–220.
- Chase MW, Palmer JD. 1989. Chloroplast DNA systematics of the lilioid monocots: feasibility, resources, and an example from the Orchidaceae. *American Journal of Botany* **76**: 1720–1730.
- Chung JM, Park KW, Park C-S, Lee S-H, Chung MG, Chung MY. 2009. Contrasting levels of genetic diversity between the historically rare orchid *Cypripedium japonicum* and the historically common *Cypripedium macranthos* in South Korea. *Botanical Journal of the Linnean Society* **160**: 119–129.
- Corriveau JL, Coleman AW. 1988. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *American Journal of Botany* **75**: 1443–1458.
- Cozzolino S, Cafasso D, Pellegrino G, Musacchio A, Widmer A. 2004. Hypervariable plastid locus variation and intron evolution in the *Anacamptis palustris* lineage. *Genome* **47**: 999–1003.
- Cozzolino S, Cafasso D, Pellegrino G, Musacchio A, Widmer A. 2007. Genetic variation in time and space: the use of herbarium specimens to reconstruct patterns of genetic variation in the endangered orchid *Anacamptis palustris*. *Conservation Genetics* **8**: 629–639.
- Darwin C. 1877. *The various contrivances by which orchids are fertilised by insects*, 2nd edn. London: J. Murray.
- Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemistry Bulletin, Botanical Society of America* **19**: 11–15.
- 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.
- Farrell L. 1999. *Cypripedium calceolus* L. (Orchidaceae). Lady's slipper orchid. In: Wigginton MJ. ed. *British Red Data Books. I. Vascular plants*. Peterborough: Joint Nature Conservancy Council, 114.
- Fay MF, Cowan RS. 2001. Plastid microsatellites in *Cypripedium calceolus* (Orchidaceae): genetic fingerprints from herbarium specimens. *Lindleyana* **16**: 151–156.
- Fay MF, Krauss SL. 2003. Orchid conservation genetics in the molecular age. In: Dixon KW, Kell SP, Barrett RL, Cribb PJ. eds. *Orchid conservation*. Kota Kinabalu: Natural History Publications (Borneo), 91–112.
- Fay MF, Qamaruz-Zaman F, Cowan RS, Thornton H. 2002. Molecular techniques for orchid conservation genetic studies. In: Clark J, Elliott WM, Tingley G, Biro J. eds. *Proceedings of the 16th World Orchid Conference*. Vancouver: Vancouver Orchid Society, 152–157.
- Fay MF, Cowan RS, Leitch IJ. 2005. The effects of genome size on the quality and utility of AFLP fingerprints. *Annals of Botany* **95**: 237–246.
- Fay MF, Smith RJ, Zuiderduin K, et al. 2007. How does hybridization influence the decision making process in conservation? The genus *Orchis* (Orchidaceae) as a case history. *Lankesteriana* **7**: 135–137.
- Forrest AD, Hollingsworth ML, Hollingsworth PM, Sydes C, Bateman RM. 2004. Population genetic structure in European populations of



- Spiranthes romanzoffiana* set in the context of other genetic studies on orchids. *Heredity* **92**: 218–227.
- Garner TWJ. 2002.** Genome size and microsatellites: the effect of nuclear size on amplification potential. *Genome* **45**: 212–215.
- Jacquemyn H, Brys R, Vandepitte K, Honnay O, Roldán-Ruiz I, Wiegand T. 2007.** A spatially explicit analysis of seedling recruitment in the terrestrial orchid *Orchis purpurea*. *New Phytologist* **176**: 448–459.
- Jersáková J, Malinová T. 2007.** Spatial aspects of seed dispersal and seedling recruitment in orchids. *New Phytologist* **176**: 237–241.
- King RA, Ferris C. 2002.** A variable minisatellite sequence in the chloroplast genome of *Sorbus* L. (Rosaceae: Maloideae). *Genome* **45**: 570–576.
- Kull T, Paaver T. 1997.** Patterns of aspartate aminotransferase variation in relation to population size, founder effect, and phylogeographic history in *Cypripedium calceolus*. *Proceedings of the Estonian Academy of Sciences. Biology and Ecology* **46**: 4–11.
- Leitch IJ, Fay MF. 2008.** Plant genome horizons: Michael Bennett's contribution to genome research. *Annals of Botany* **101**: 737–746.
- Machon N, Bardin P, Mazer SJ, Moret J, Godelle B, Austerlitz F. 2003.** Relationship between genetic structure and seed and pollen dispersal in the endangered orchid *Spiranthes spiralis*. *New Phytologist* **157**: 677–687.
- Nelson EC. 1994.** Historical data from specimens in the herbarium, National Botanic Gardens, Glasnevin, Dublin (DBN), especially on *Cypripedium calceolus*. *BSBI News* **67**: 21–22.
- Oxelman B, Liden M, Berglund D. 1997.** Chloroplast *rps16* intron phylogeny of the tribe Sileneae (Caryophyllaceae). *Plant Systematics and Evolution* **206**: 393–410.
- Petit RJ, El Mousadik A, Pons O. 1998.** Identifying populations for conservation on the basis of genetic markers. *Conservation Biology* **12**: 844–855.
- Pillon Y, Fay MF, Hedrén M, et al. 2007.** Evolution and temporal diversification of western European polyploid species complexes in *Dactylorhiza* (Orchidaceae). *Taxon* **56**: 1185–1208.
- Pons O, Petit RJ. 1996.** Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* **144**: 1237–1245.
- Posada D, Crandall KA. 2001.** Intraspecific gene genealogies: grafting trees into networks. *Trends in Ecology and Evolution* **16**: 37–45.
- Qamaruz-Zaman F. 2000.** *Conservation genetics of rare and endangered British orchids*. PhD thesis, University of Cambridge, UK.
- Qamaruz-Zaman F, Fay MF, Parker JS, Chase MW. 1998.** Molecular techniques employed in the assessment of genetic diversity: a review focusing on orchid conservation. *Lindleyana* **13**: 259–283.
- Schneider S, Roessli D, Excoffier L. 2000.** *Arlequin: a software for population genetics data analysis*, Version 2.000. Geneva: Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva.
- Shipunov AB, Fay MF, Pillon Y, Bateman RM, Chase MW. 2004.** *Dactylorhiza* (Orchidaceae) in European Russia: combined molecular and morphological analysis. *American Journal of Botany* **91**: 1419–1426.
- Shipunov AB, Fay MF, Chase MW. 2005.** Taxonomic status of *Dactylorhiza baltica* (Orchidaceae) from European Russia: evidence from molecular markers and morphology. *Botanical Journal of the Linnean Society* **147**: 257–274.
- Small RL, Ryburn JA, Cronn RC, Seelanan T, Wendel JF. 1998.** The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. *American Journal of Botany* **85**: 1301–1315.
- Summerhayes VS. 1951.** *Wild orchids of Britain*. London: Collins.
- Swofford D. 2002.** *PAUP\* 4.0b10: Phylogenetic analysis using parsimony (\* and other methods)*. Sunderland, MA: Sinauer Associates.
- 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.
- Vos P, Hogers R, Bleeker M, et al. 1995.** AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**: 4407–4414.
- Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. 1990.** DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* **18**: 6231–6235.