

# A population genetic analysis of chloroplast DNA in *Phacelia*

FOSTER LEVY\*, JANIS ANTONOVICS†, JOHN E. BOYNTON† & NICHOLAS W. GILLHAM‡

Department of Biological Sciences, East Tennessee State University, Johnson City, TN 37614 and Departments of †Botany and ‡Zoology, Duke University, Durham, NC 27708, U.S.A.

Hierarchical sampling from populations, incipient and recognized varieties within *Phacelia dubia* and *P. maculata* has revealed high levels of intraspecific polymorphism in chloroplast DNA. Much of the variation is partitioned between populations as evidenced by population-specific variants at fixation in all three populations of *P. dubia* var. *interior* and in both populations of *P. maculata*. Nine of 16 populations were polymorphic for cpDNA haplotypes. A total of 16 haplotypes was found in a sample of 106 individuals; the most common occurred in eight of the 16 populations and in 31 per cent of the individuals in the entire sample. A phylogenetic analysis revealed four basic plastome types. The two major groups of plastomes were separated by four independent base-pair mutations which suggests an ancient split in the evolution of plastid genomes. Representatives from each major plastome division were found in each of five populations spanning two allopatric varieties of *P. dubia*. The geographical distribution of haplotypes and lack of evidence for recent admixture argue against migration as a source of the polymorphism. It is more likely that the current taxonomic varieties are descendants of a polymorphic common ancestor.

**Keywords:** chloroplast DNA, genetic diversity, haplotype, intraspecific polymorphisms, population structure.

## Introduction

Intraspecific variation in chloroplast DNA (cpDNA) has now been widely documented (Harris & Ingram, 1991; Soltis *et al.*, 1992). Population analyses based upon the chloroplast genome are beginning to be applied to problems in intraspecific phylogeography in plants (Soltis *et al.*, 1992; Hooglander *et al.*, 1993; Mayer & Soltis, 1994) in ways analogous to the use of mitochondrial DNA (mtDNA) in studies of animals (Avise *et al.*, 1987). In addition, data from cpDNA have been used to examine population structure within species (Neale *et al.*, 1988; Milligan, 1991; Soltis *et al.*, 1992; McCauley, 1994; Petit *et al.*, 1994), haplotype diversity within and between populations (Milligan, 1991; Hong *et al.*, 1993; Byrne & Moran, 1994), and to search for correlations between chloroplast variation and morphological (Terauchi *et al.*, 1991) or breeding system (Wolff & Schaal, 1992) characteristics.

Many detailed examinations of intraspecific variation in cpDNA concern agricultural crops. The goals

of these studies have been to outline the amount, geographical location and genetic nature of variation, often by comparing cultivars and wild relatives (Kochert *et al.*, 1991). However, population structure is best revealed via a hierarchical sampling design, e.g. individuals within populations and populations within species.

Widespread occurrence of intraspecific variation in cpDNA has opened the door for studies designed to infer mechanisms leading to population differentiation. Recent studies have used hierarchical sampling to reveal migration patterns that led to the recolonization of glaciated habitats in oaks (Petit *et al.*, 1994), to provide evidence for isolation by distance among populations of *Eucalyptus* (Byrne & Moran, 1994) and to show that populations of *Trifolium pratense* appear to approach panmixia (Milligan, 1991).

The goal of this study was to analyse the population structure of a subspecifically differentiated species and its nearest relative. The species, *Phacelia dubia* (L.) Trel., consists of three allopatric varieties and two incipient varieties, all of which differ morphologically and which show patterns of infertile-

\*Correspondence.

ity when intercrossed (Levy, 1991a,b). More specifically, we used data on population structure and haplotype diversity to infer mechanisms underlying plastome differentiation and to suggest possible forces responsible for observed patterns. We examined evidence for gene flow among populations and whether the distribution of variants may best be attributed to random genetic drift, selection, and/or historical factors such as random sorting of plastome lineages during taxon divergence.

The study has uncovered some unexpected patterns. *P. dubia* harbours a surprisingly high level of genetic diversity within each level (varieties, incipient varieties, populations) of the intraspecific hierarchy. There was evidence for an ancient plastome polymorphism whose current representatives coexist within several allopatric populations. A phylogeny of plastid haplotypes and their distribution within taxa indicated taxon divergence probably proceeded from a polymorphic ancestor.

## Materials and methods

### Study organism

*Phacelia dubia* was chosen to examine the distribution of intraspecific variation in cpDNA because populations are spatially discrete and subspecific varieties express a broad range of geographical, ecological and morphological variation (Murdy, 1966; Levy, 1991a). The varieties within *P. dubia* constitute a monophyletic group which exhibits complete reproductive isolation from *P. maculata*, their nearest relative (Gillett, 1964; Murdy, 1966). Reproductive isolation among the subspecific taxa within *P. dubia* is incomplete.

The three allopatric, formally recognized taxonomic varieties of *P. dubia* are *P. dubia* var. *dubia* (= *dubia*), *P. dubia* var. *georgiana* McVaugh (= *georgiana*) and *P. dubia* var. *interior* Fernald (= *interior*). Based upon morphological, geographical and reproductive relationships, *dubia* has been hypothesized as the progenitor of the other two varieties (Murdy, 1966; Levy, 1991a,b).

Two incipient varieties were also included in this study to determine (a) the extent of intrapopulation polymorphisms, (b) levels of differentiation from recognized varieties, and (c) their phylogenetic relationships to recognized varieties. The incipient varieties were designated *P. dubia* 'imitator' (= 'imitator') and *P. dubia* 'railroad' (= 'railroad'). Populations of 'imitator' are partially reproductively isolated from both *dubia* and *georgiana* (Levy, 1991b) and they are located in a region (South

Carolina) between the known ranges of the recognized varieties *dubia* and *georgiana*. 'Railroad' consisted of two large populations in southwestern Virginia, within the range of and in proximity to *dubia* populations (Levy, 1991a). Although partially reproductively isolated from *dubia* by a nuclear-cytoplasmic barrier, these plants did not differ morphologically from *dubia* (Levy, 1991a).

Plants for cpDNA extraction were grown from seeds which were collected in the field in 1987 and 1988. All plants were grown in the Duke University Phytotron. Chloroplast DNA was analysed in plants from four populations of *P. maculata* and from two to four populations within each of the five varieties of *P. dubia* (three recognized varieties: *dubia*, *interior*, and *georgiana*; two incipient varieties: 'railroad' and 'imitator'). Population designations correspond to those in Levy (1991a) with the addition of g-4, Douglas Co., GA (*georgiana*); m-3, Edgefield Co., SC (*P. maculata*); m-4, Gwynette Co., GA (*P. maculata*).

### Chloroplast DNA isolation

The methods of cpDNA extraction and purification were similar to those described in Palmer (1982) and Kemble (1987). Chloroplast DNAs that produced partial banding patterns suggestive of incomplete digestion were further purified in a 1 mL column of 100–200 mesh agarose beads (Bio-Rad A-50) followed by ethanol precipitation.

### Restriction enzyme analysis

Restriction fragments were separated on 0.8 per cent agarose gels at 50 mA current for 12 to 16 h in 1×TBE buffer. The gels were stained in 1 mg/L ethidium bromide for 30 min, destained in three changes of distilled water for 30 min each, and photographed on Kodak Tri-X film with a Polaroid camera. Banding patterns were analysed visually.

To facilitate visual scoring, a preliminary survey conducted on a small subset of cpDNAs was carried out to identify restriction enzymes that (a) generated banding patterns free of partially digested fragments and (b) produced a moderate (15 to 30) number of fragments whose bands were evenly distributed throughout the gel. Twenty-two enzymes were tested; 18 with six base-pair recognition sequences (*Apa*I, *Bam*HI, *Bcl*II, *Bgl*I, *Bgl*II, *Bst*EI, *Cla*I, *Eco*RI, *Eco*RV, *Hind*III, *Kpn*I, *Pst*I, *Pvu*II, *Sma*I, *Sst*I, *Sst*II, *Xba*I, *Xho*I), one with a five base-pair recognition sequence (*Ava*II), and three with four base-pair recognition sequences (*Hae*III, *Hha*I, *Hpa*II). Five of

these (*Bam*HI, *Bgl*II, *Hind*III, *Sst*I, *Xba*I) were used to estimate genome size and length of the inverted repeat on each of four cpDNAs. Using six restriction enzymes (*Ava*II, *Bam*HI, *Bgl*II, *Eco*RV, *Hha*I, *Hind*III), plastome type was determined for all 122 cpDNAs and complete haplotypes were obtained for 106 of the cpDNAs.

#### Scoring variation

Base-pair substitutions were inferred when the combined length of two fragments present in one cpDNA sample but absent in another, summed to the length of a newly visible fragment in the second sample. This pattern would correspond to a difference of a single restriction site between the cpDNA in the two lanes. Several suspected base-pair substitutions were confirmed by filter hybridizations. When a fragment was either gained or lost, but the pattern corresponding to a simple base-pair substitution was not evident, this pattern was scored as a presence/absence variant. Presence/absence variants may have arisen by length mutations such as insertions or deletions but some were probably also the result of base-pair substitutions in which one cpDNA fragment was too small to detect.

#### Partial mapping of variant fragments

A partial mapping procedure was used to distinguish whether the changes associated with different variants were located in physical proximity to each other on the genome. Two cpDNA samples may differ by several RFLPs, but if these variants map to a similar region of the genome it may indicate that they have all arisen by a single mutation event, such as a large insertion or deletion. Five plastome-specific variants were partially mapped. In this procedure the largest variant fragment in a particular enzyme digest was electroeluted either in dialysis tubing (slices of the gels containing cpDNA fragments were electrophoresed in  $0.5 \times$  TBE buffer for 1 h) or in an electroeluter apparatus (IBI Model UEA) followed by purification (Elutip D columns, Schlichter & Schull) and ethanol precipitation. The recovered fragments were nick translated (BRL Nick Translation System no. 8160SB) with  $^{32}$ P and used as probes. These radioactively labelled probes were hybridized to nitrocellulose filters onto which samples of cpDNAs encompassing several different variants were blotted following digestion of each with a panel of six restriction enzymes.

The following strategy was used to determine whether different variants resulted from indepen-

dent mutations and whether they mapped to neighbouring regions of the genome (as expected when length mutations cause the simultaneous loss or gain of nearby restriction sites). Because each probe was derived from a variant fragment and each probe was hybridized to samples digested with the panel of six restriction enzymes, variation was expected among cpDNAs digested with the restriction enzyme giving rise to the probe fragment. If hybridization of a particular probe also revealed variation in cpDNAs digested with restriction enzymes other than the one used to generate the probe, then the two variants must be in close physical proximity in the genome; that is, both variants were either encompassed by or adjoined DNA corresponding to the probe.

#### Population structure

A hierarchical sampling design (individuals within populations; populations within varieties; varieties within species) was employed to facilitate an analysis of population structure. Between-population components of haplotype diversity were estimated separately for each of the three recognized varieties of *P. dubia* as well as for the entire *P. dubia* complex of populations (including populations of the two putative varieties). The measure of population structure, theta ( $\theta$ ), was estimated from a haploid model as outlined by Weir (1990). A mean value for theta and its standard deviation were computed by jackknifing over populations using a computer program written by B. Weir.

#### Phylogenetic inference

Phylogenetic relationships among plastid haplotypes were assessed in a parsimony analysis using unweighted characters in the branch and bound algorithm of PAUP (Swofford, 1991). The plastomes in *P. maculata* served as an outgroup for plastomes in *P. dubia*. A goal of this study was to determine whether intraspecific polymorphism occurred in cpDNA and if so, how the polymorphism may influence phylogenetic inferences. The chloroplast genome has been believed to evolve more slowly than nuclear DNA (Wolfe *et al.*, 1987) and to exhibit little intraspecific variation (Palmer, 1987; Palmer *et al.*, 1988). Consequently, in many phylogenetic studies using cpDNA, only one or a few plants of a species have been considered adequate as representatives of a species. Two phylogenetic analyses were performed to investigate potential effects of limited intrataxon sampling. First, the plant with the most frequent cpDNA haplotype within each recognized

*Phacelia* taxon (*dubia*, *georgiana*, *interior*, *P. maculata*) was chosen as a representative. For example, *dubia* was represented by haplotype D-4 (Table 4); in *georgiana*, 17 of the 27 individuals carried plastome G and haplotype G-3 was chosen because it was most common within plastome G. In an alternative analysis, *georgiana* was represented by a plant carrying the most frequent haplotype in that variety regardless of the plastome type (haplotype D-1; see Table 4).

## Results

### Genome organization

The genome size was estimated as 129.9 kb and the length of the inverted repeat was estimated as 20.0 kb, both of which are within the range of most flowering plants (Palmer, 1985).

Each of 122 cpDNA samples was digested with six restriction endonucleases; the four enzymes with six base-pair recognition sequences (*Bam*HI, *Bgl*II, *Eco*RV, *Hind*III) produced a total of 100 scorable restriction fragments, the five-base cutter (*Ava*II) generated 32 scorable fragments, and the four-base cutter (*Hha*I) generated 31 scorable fragments. Approximately 884 bp were sampled which represents 0.7 per cent of the 129.9 kb *Phacelia* chloroplast genome.

### Plastome variation

Each of the 122 cpDNAs could be characterized as possessing one of four basic restriction patterns; each of the four patterns is characterized by a unique suite of restriction fragments. Differences among these four plastid types (subsequently referred to as plastomes and designated by the upper case letters 'D', 'I', 'G', and 'M') were observed in digestions with each of the six enzymes analysed. The thirteen variants that distinguished the four plastomes, and the nature of these variants are presented in Table 1. Plastome variants observed following digestion of cpDNAs with *Ava*II are illustrated in Fig. 1.

A set of four base-pair substitutions (Table 1; nos 1, 2, 4, 5) and one presumed length mutation (Table 1; no. 3) divides the entire sample into two major groups, one encompassing plastomes D and I and the other consisting of plastomes G and M. Within the G-M group, plastome M carried five additional variants that distinguished it from all other plastomes (three base-pair substitutions, two presence/absence bands) (Table 1; nos 7-11); there was one

variant (Table 1; no. 6) restricted to plastome G. In addition to the five variants shared with plastome D, plastome I differed from plastome D by two presence/absence variants (Table 1; nos 12, 13). Radioactive probes derived from cpDNA of five plastome-specific fragments were sequentially hybridized to Southern blots of cpDNA digests produced by the same enzyme giving rise to the respective variant fragments. Four of the probe DNAs distinguished the D-I from the G-M lineage (Table 1; nos 1, 2, 4, 5) and one (Table 1; no. 9) separated plastome M from the other three plastomes. In four cases, the observed variation could be attributed to the loss or gain of a restriction site, presumably by a single base-pair substitution. These five fragments encompassed 33.4 kb of the chloroplast genome (*Ava*II, 9.0 kb and 2.8 kb; *Eco*RV, 2.6 kb; *Hha*I, 13.5 kb; *Hind*III, 5.5 kb).

For each hybridized probe, the banding patterns of representatives of each plastome were identical except in hybridizations to digests by the enzyme which generated the probe fragment. These hybridizations indicated that variant fragments produced by different restriction enzymes must have been located in different regions of the chloroplast genome. Hence, given the extent of the region used as probe DNA (33.4 kb), the location of variant fragments in different regions of the genome, the lack of apparent major differences in genome size among these cpDNAs, and the evidence that four of these variants were resulted from base-pair substitutions, it is unlikely that differences among the four plastomes resulted from a single rearrangement but rather that they differed by several independent point mutations and possible length mutations.

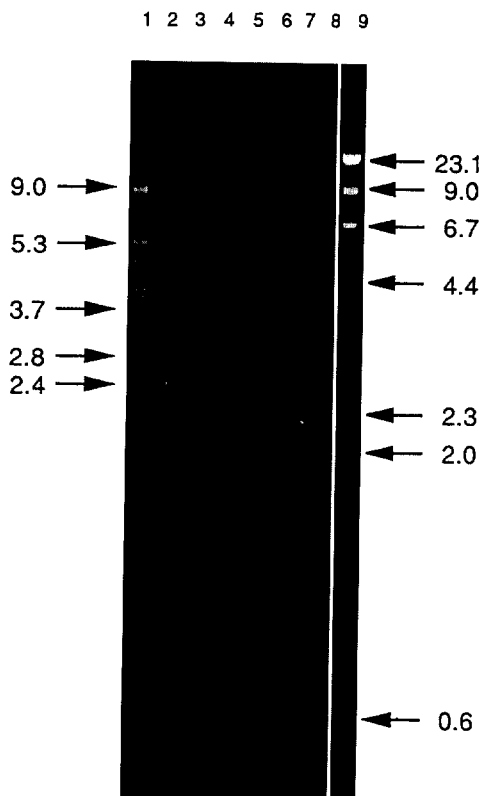
### Distribution of the plastomes among taxa

The distribution and frequencies of the four plastomes in the various taxa (Table 2) shows that plastome M was unique to *P. maculata* and all plants in that taxon carried plastome M. Plastomes D, G and I were restricted to *P. dubia*; plastome I was restricted to and fixed within populations of *interior*; plastome D was fixed in the two putative varieties, 'railroad' and 'imitator'. Plastome D was found in all but one of the 28 representatives of *dubia* and it was present at varying frequencies in all four populations of *georgiana* (Table 2). Plastome G was represented in samples from all populations of *georgiana* (Table 2). The wide range of plastome frequencies among *georgiana* populations may reflect variation arising from small sample sizes, but the presence of both plastomes D and G within all sampled *georgiana*

**Table 1** Variants characteristic of the four *Phacelia* plastomes

Code no.	Enzyme	Plastome				Variant
		M	G	D	I	
1	<i>Ava</i> II*	—	—	+	+	5.3+3.7→9.0
2	<i>Ava</i> II*	—	—	+	+	2.8→2.4+0.4
3	<i>Bam</i> HI	—	—	+	+	3.9→2.6
4	<i>Eco</i> RV*	—	—	+	+	2.6→2.1+0.5
5	<i>Hind</i> III*	—	—	+	+	5.5→5.2+0.3
6	<i>Hha</i> I	—	+	—	—	3.8→2.2+1.6
7	<i>Ava</i> II	—	+	+	+	4.0→4.3
8	<i>Eco</i> RV	—	+	+	+	14.0+9.0→11.0+12.0
9	<i>Hha</i> I*	—	+	+	+	12.0+1.5→13.5
10	<i>Hha</i> I	—	+	+	+	9.8→8.0+1.8
11	<i>Hind</i> III	—	+	+	+	—2.5
12	<i>Ava</i> II	—	—	—	+	4.2→4.3
13	<i>Bgl</i> II	—	—	—	+	+2.0

The changes indicated under 'variant' refer to differences between plastome M and other plastomes. A plus indicates that the change listed under 'variant' occurs in a particular plastome. Single entries under 'variant' signify presence or absence of fragments not attributable to restriction site changes or obvious length mutations. DNA fragment lengths are in kilobase pairs. \*Variants confirmed by Southern analysis.



populations (Table 2) suggests that both are common throughout the range of *georgiana*.

#### Taxonomic and geographical distribution of haplotypes

Complete haplotypes were determined for 106 of the 122 samples. In addition to the 13 diagnostic plastome variants, 16 additional variant banding patterns (referred to as nonplastome variants and designated by the lower case letters a–q) were detected (Table 3). One pair of these (a, e) co-occurred in 26 cpDNA samples from plants representing members of five populations that spanned three taxonomic varieties (haplotypes D-2, D-3, D-4 in Table 4). A second pair (f, g) was only found in a single plant and to provide conservative estimates of

**Fig. 1** Plastome-related variation observed following *Ava*II digestion. Chloroplast DNAs from individual plants from population d-1 (lanes 1–3), g-3 (lanes 4–6) and g-1 (lanes 7–8). Plastome D has a 9.0 kb fragment (lanes 1, 2, 3, 7) which, in plastome G (lanes 4, 5, 6, 8) has an internal *Ava*II recognition site splitting the 9.0 kb fragment into 5.3 kb and 3.7 kb fragments. Plastome G has a 2.8 kb fragment which is split into 2.4 and 0.4 fragments in plastome D. Note the intrapopulational plastome polymorphism in g-1 (lanes 7, 8). Size markers are in lane 9.

**Table 2** Frequencies of the four plastomes in *Phacelia* taxa and the frequencies of these plastomes in populations polymorphic for two plastomes

Taxon	N	Plastome			
		D	G	M	I
<i>P. dubia</i> var. <i>dubia</i>	28	0.96	0.04	—	—
<i>P. dubia</i> var. <i>interior</i>	21	—	—	—	1.0
<i>P. dubia</i> var. <i>georgiana</i>	29	0.34	0.66	—	—
<i>P. dubia</i> var. 'railroad'	15	1.0	—	—	—
<i>P. dubia</i> var. 'imitator'	16	1.0	—	—	—
<i>P. maculata</i>	13	—	—	1.0	—
Polymorphic populations					
<i>dubia</i> d-2	8	0.87	0.13	—	—
<i>georgiana</i> g-1	9	0.22	0.78	—	—
<i>georgiana</i> g-2	9	0.67	0.33	—	—
<i>georgiana</i> g-3	9	0.11	0.89	—	—
<i>georgiana</i> g-4	2	0.50	0.50	—	—

N = number of individuals sampled.

**Table 3** Variation within plastomes in *Phacelia*

Enzyme	Variant code	Variant	Plastome
<i>Ava</i> II	a	2.6→2.5	D
<i>Ava</i> II	b	2.0→1.6	I
<i>Bam</i> HI	c	4.6→2.4+2.2	G
<i>Bgl</i> II	d	+3.8	G
<i>Bgl</i> II	e	3.4+1.9→1.3	D
<i>Bgl</i> II	f	8.2+4.2→15.0	G
<i>Eco</i> RV	g	5.8+2.0→7.8	G
<i>Eco</i> RV	h	2.3→2.2	G
<i>Eco</i> RV	i	+0.9	I
<i>Eco</i> RV	j	1.8→1.0+0.8	D, I
<i>Hha</i> I	k	2.6→2.3	I
<i>Hha</i> I	l	-5.0	D, G
<i>Hha</i> I	m	-2.0	D
<i>Hha</i> I	n	3.0→2.5	G
<i>Hha</i> I	p	3.0→3.2	D, I
<i>Hind</i> III	q	2.6→2.7	D, M

These variants occur in addition to the respective diagnostic variants of each plastome. Single entries under 'variant' signify presence or absence of fragments not attributable to restriction site changes or obvious length mutations. DNA fragment lengths are given in kilobase pairs.

divergence in phylogenetic analyses, was also assumed to be associated. No other pairs of variant fragments showed obvious associations. Consequently, each of the 14 remaining variants and the two associated pairs were considered to have arisen

by independent mutations. Sixteen different haplotypes were distinguished (designated by the upper case plastome letter code and a genotype number): six were variants of the D plastome, four were variants of the I plastome, there were four variants within the G plastome and two variants within the M plastome (Tables 3, 4; Fig. 3a).

In several instances, identical haplotypes (for example, haplotypes D-1 and G-2) occurred in individual plants collected from two different taxonomic varieties (Table 4). One population within *interior* (i-2) and one of *P. maculata* (m-3) were fixed for unique variants, that is, all members of these populations carried a variant not encountered in any other population. In a *georgiana* population (g-3), which was polymorphic for plastomes D and G, all seven individuals with plastome G carried a unique variant (Table 4).

#### Population structure

Nine of the 16 populations surveyed were polymorphic for chloroplast DNA haplotypes (Table 4). Five populations (one of the three *dubia* populations (d-1) and all four *georgiana* populations) contained members with different plastomes; plastomes D and G were represented in each of these five populations (Table 2). Six populations showed variation within a particular plastome and in three populations, there was a single individual with a unique variant (Table 4).

Considering the entire sample, 10 of the 16 haplotypes were rare, that is, they occurred with a frequency of less than 0.05 (Fig. 2). One haplotype (D-1) was by far the most common (frequency = 0.31) and this haplotype was also the most widely distributed, occurring in eight populations that spanned two recognized and two incipient varieties of *P. dubia* (Table 4). In contrast, most (11 of 16) haplotypes were restricted to single populations (Fig. 2).

In *P. dubia*, most of the haplotype diversity was distributed between populations ( $\theta = 0.61$ ; SD = 0.11). However, the three recognized varieties showed very different levels of substructuring. Substructuring was greatest in var. *interior* where all haplotypes were population-specific ( $\theta = 0.86$ ; SD = 0.15). Substructuring was lowest within var. *dubia* where a single haplotype (D-4) was the most frequent within each of the three populations ( $\theta = 0.19$ ; SD = 0.24). The level of substructuring within var. *georgiana* was intermediate between the other two varieties ( $\theta = 0.45$ ; SD = 0.24) which reflects a combination of unique haplotypes (G-3,

Table 4 Distribution of cpDNA haplotypes in populations of *Phacelia*

Haplotype	Taxon														Haplotype totals				
	<i>dubia</i>				<i>georgiana</i>				<i>interior</i>				<i>P. maculata</i>		No. of pops.	No. of indiv.			
	d-1	d-2	d-3	d-4	g-1	g-2	g-3	g-4	i-1	i-2	i-3	r-1	r-2	p-1			p-2	m-1	m-2
D-1	1	3	2	2	6	1					7	6	7					8	33
D-2	2																	2	9
D-3	2																	1	2
D-4	3	3	8			1												4	15
D-5											1							1	1
D-6														2				1	2
I-1																			
I-2										2								1	2
I-3									7									1	7
I-4									5									1	5
G-1									1									1	1
G-2								3	1									3	5
G-3	1							4										2	5
G-4								1										1	7
M-1																		1	1
M-2																7		1	7
																	4	1	4

Entries refer to the number of individuals of each haplotype.

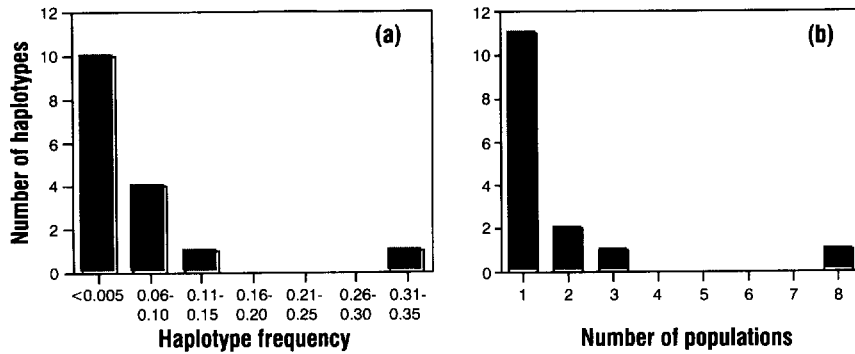


Fig. 2 (a) Haplotype frequencies in the entire sample from *Phacelia dubia*. (b). Frequency distribution of haplotypes among populations of *P. dubia*.

G-4) and a haplotype (D-1) that was shared among three of the four populations (Table 4).

#### Phylogenetic analyses

The parsimony analysis resulted in one shortest tree which required 33 steps to assign 29 characters (consistency index = 0.88) (Fig. 3a). Four characters showed homoplasy; three (l, p, q) resulted from parallelisms, one (j) was a reversal. A suite of five mutations (Table 1, nos 1–5) separated plastomes D and I from plastomes G and M. Five additional variants characterized plastome M. Haplotype-specific apomorphies distinguished lineages within plastome types (Fig. 3a).

The analysis using a single plant to represent each taxon (using the most frequent haplotype of the most common plastome within each of the taxonomically recognized taxa) indicated a sister group relationship between *georgiana* and *P. maculata*. Thus, the taxon *P. dubia* would be considered paraphyletic. The resultant tree of length 23 (six of 29 characters were constant) showed no homoplasy (Fig. 3b). Estimates of relatedness were drastically altered when the individual chosen to represent *georgiana* was one carrying the most frequent haplotype (regardless of plastome type) in that variety, haplotype D-1. In this analysis, a lack of phylogenetically informative characters resulted in an unresolved polytomy (Fig. 3c). This simple comparison shows that tree topology and distances can differ greatly depending upon the sampling scheme for individuals within taxa.

## Discussion

#### Intraspecific polymorphism

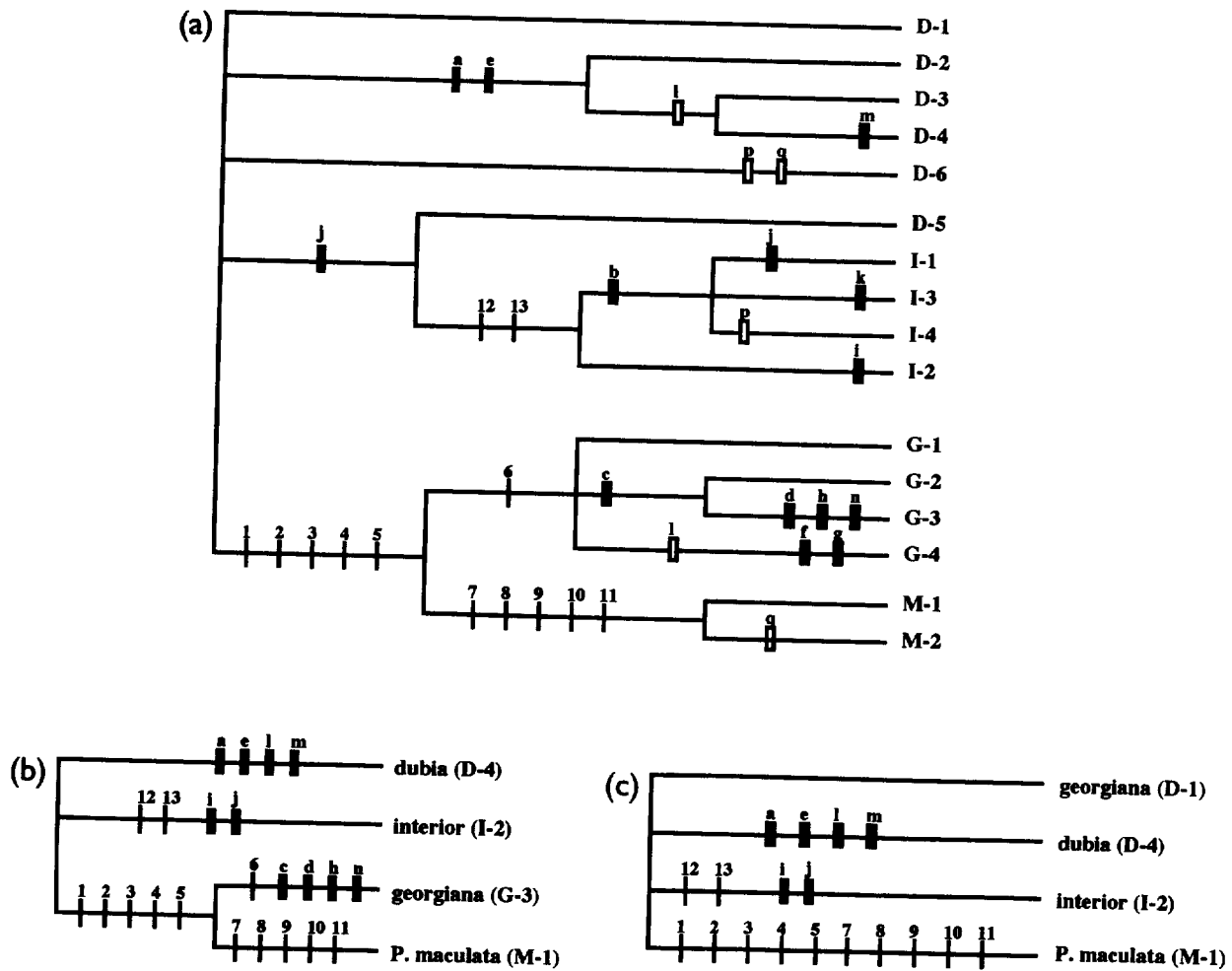
Until recently, it has been an unwritten dogma that cpDNA shows little intraspecific variation, even though the earliest surveys of cpDNA variation in natural populations detected intraspecific variation

(Scowcraft, 1979; Palmer & Zamir, 1982). More recent and thorough analyses of intraspecific cpDNA variation suggest that polymorphism is common, but that the level of variation depends upon the particular species (Hong *et al.*, 1993). In the present study, extensive variation was found within *P. dubia* and *P. maculata*. Nine of 16 populations were polymorphic for cpDNA haplotypes (Table 4). The relatively high level of interpopulation polymorphism in *Phacelia* was comparable to levels observed in *Hordeum vulgare* (Neale *et al.*, 1988; 10 of 19 populations polymorphic) and *Eucalyptus nitens* (Bryne & Moran, 1994) but it was not as high as levels observed in *Trifolium pratense* (Milligan, 1991). Lower levels of polymorphism characterized European oaks (Petit *et al.*, 1994) and *Lupinus texensis* (Banks & Birky, 1985).

A large proportion of the cpDNA diversity in *P. dubia* occurred as between-population differences ( $\theta = 0.61$ ). A similar pattern of relatively high genetic diversity primarily partitioned between populations was also observed in *H. vulgare* where  $\theta = 0.67$  (Milligan, 1991, from data in Neale, 1988) and in bishop pine ( $\theta > 0.80$ ; Hong *et al.*, 1993). In contrast, extremely low between-population differentiation ( $\theta = 0.02$ ) for cpDNA haplotypes was found in the highly polymorphic *T. pratense*, a pattern which is suggestive of panmixia (Milligan, 1991).

Haplotype frequency distributions also vary greatly among species: a single haplotype dominated the *Lupinus* survey (Banks & Birky, 1985), occurring in 88 per cent of the 100 individuals surveyed; in the current study, the most common of the 16 *Phacelia* haplotypes was found in 31 per cent of the 106 individuals haplotyped, and in a sample of 245 plants of *Hordeum*, each of the three haplotypes comprised 30–40 per cent of the sample (Neale *et al.*, 1988). The most frequent haplotype in *Eucalyptus* occurred in only 19 per cent of the individuals (Bryne & Moran, 1994). Haplotype frequencies





**Fig. 3** Cladograms showing relationships (a) among haplotypes; (b) among recognized taxa using the single most common cpDNA haplotype in each taxon to represent that taxon; (c) same as (b) but replacing the *georgiana* individual with one carrying the D plasmome that occurred in *georgiana* populations. Variants that show homoplasy are shown as open (parallelisms) and gray-shaded (reversal) rectangles. Variant codes correspond to those in Table 1 (numeric) and Table 3 (alphabetic).

varied greatly among very local populations of *Silene alba* (McCauley, 1994). Plastome variants that were unique to single populations and fixed within those populations were found in two populations each of *P. dubia* and *P. maculata*, also in two populations of *Eucalyptus* (Bryne & Moran, 1994), and in one population of *Lupinus* (Banks & Birky, 1985). In *Hordeum*, several populations lacked cpDNA haplotypes that were common in populations less than 30 km distant (Neale *et al.*, 1988).

Do data from the limited number of detailed intraspecific investigations suggest generalizations regarding the amount and partitioning of cpDNA variation within species? The most noteworthy

conclusion must be that intraspecific polymorphism can be found in most species for which extensive samples have been analysed. Species differ in the number of haplotypes and their partitioning within and between populations as demonstrated by differences among the closely related closed-cone pines in California (Hong *et al.*, 1993). The amount of genetic difference between haplotypes within species may be small as is often observed, but in *P. dubia* and *Coreopsis grandiflora* (Mason *et al.*, 1994) haplotypes differed by as many as 14 and 19 mutations, respectively. The species-specific nature of cpDNA variation should not be surprising; levels of polymorphism depend upon the mutation rate and effective

population size. Although the chloroplast genome is thought to evolve slowly, little is known about its mutation rate other than that there is an apparently higher rate for length variants compared to point mutations (Hong *et al.*, 1993). In contrast, extensive data exist that show that species differ in all components (e.g. census size, mating system, migration rate, etc.) of effective population size. In fact, the species-specific nature of cpDNA variation is similar to that found in analyses of nuclear-encoded allozyme loci (Hamrick *et al.*, 1979). Species specificity may preclude generalizations but it does facilitate the use of cpDNA to infer population processes as demonstrated below.

#### *Geographical patterns and genotypic distributions*

Avise *et al.* (1987) distinguished five distributional patterns of mtDNA in populations. In the most common (Category I), phylogenetic occurrence of genomes was correlated with geographical separation of these genomes. The restriction of *Phacelia* cpDNA plastomes I and M to populations of *interior* and *M. maculata*, respectively, exemplify this pattern and it is analogous to the geographical partitioning of mtDNA haplotypes into isolated river drainages in several species of fish (Bermingham & Avise, 1986).

The observation of identical cpDNA haplotypes occurring as polymorphisms within several geographically isolated populations is an unusual pattern (Category II of Avise *et al.*, 1987). In the warmouth sunfish (*Lepomis gulosus*), Bermingham & Avise (1986) found two individuals in the western region of the range that had mtDNA typical of eastern populations. In plants, coexistence of individuals with highly divergent chloroplast plastomes was recently reported in *Coreopsis grandiflora* (Mason *et al.*, 1994). If we consider the possibility that, at least for *Phacelia*, intraspecific polymorphism in cpDNA is the rule rather than an exception, then several phenomena must be accounted for; the relative time and taxonomic origin of the major plastome types, the current distribution pattern of haplotypes within and between populations and the possible causes of monomorphism within populations.

Plastome differences in *Phacelia* were caused by several apparently independent base pair mutations. In light of the relatively low evolutionary rate for cpDNA, and assuming a similar rate among these closely related lineages, the number of mutations separating plastomes strongly suggests a relatively ancient origin for the D/I–G/M plastome split.

Co-occurrence of the highly divergent D and G plastomes in different reproductively isolated varieties of *P. dubia* suggests a long-standing polymorphism or recent migration. If the derivation of plastome types was indeed ancient, we must enquire why several populations remain polymorphic. Three alternatives should be considered: polymorphisms within populations may result from recent migration, they may be actively maintained by selection, or they may be neutral. If neutral, polymorphisms are expected to be transient and the frequency of a specific haplotype would reflect its point in a particular trajectory of genetic drift. Drift trajectories within populations are independent and we would not expect identical haplotypes to have similar frequencies in different populations. Although sample sizes within populations are small, inspection of the data shows that frequencies of individual haplotypes differ greatly among varieties and among populations within varieties (Table 4). The drift model cannot be ruled out.

*Migration.* All four haplotypes in *interior* and both haplotypes in *P. maculata* were population-specific and they were each fixed within their respective populations. The uniqueness of each of these populations is maintained in spite of physical proximity to other populations; within the respective taxa, populations were only 10–30 km distant from other sampled populations. Furthermore, the 'private' (i.e. restricted to a single population; Slatkin, 1985) haplotypes represent a wide spectrum of population frequencies (Table 4). If these private haplotypes represented recent mutations, most would be expected to occur at low frequencies within their respective populations. Using Slatkin's (eqn 3 in Slatkin, 1985) method based upon rare alleles, the estimate of  $Nm = 0.03$  ( $N$  = population size;  $m$  = migration rate) for all populations within *P. dubia* is far below the value of one; it implies that migration is not sufficient to prevent population differentiation, a conclusion that is consistent with the finding that most haplotype variation occurred between populations ( $\theta = 0.61$ ). When  $Nm$  is estimated from  $q$  (using eqn 3 in Ennos, 1994), the estimate of  $Nm = 0.16$  for all populations within *P. dubia* is also sufficiently low to expect subdivision. Most strikingly, patterns in *Phacelia* are quite different from those in *Trifolium* where only 3 per cent of the haplotype diversity occurred between populations and  $Nm = 5.9$  was consistent with higher levels of gene flow (Milligan, 1991).

Although the possibility of historical introgression can never be eliminated, morphological patterns, reproductive relationships and analysis of nuclear-

encoded allozymes provide further evidence that recent migration is an unlikely cause of the D–G plastome polymorphism (Murphy, 1966; Levy, 1991a,b). More likely explanations for the distribution of polymorphisms in *P. dubia* should be based upon historical factors such as a relatively recent derivation of *dubia* and *georgiana* from a polymorphic common ancestor or hypotheses that postulate selective differences among plastomes.

**Selection.** Selective differences among plastomes may be hypothesized to account for the widespread plastome polymorphism: for example, one plastome may be favoured over the other in different microhabitats or seasons. Is it plausible that selective differences among organelle genomes underlie levels of polymorphism? There are no data that show fitness differences among plants that differ in chloroplast genotypes but recent analyses of mitochondrial DNA sequence diversity in *Drosophila* are not consistent with neutrality; the observed patterns are more suggestive of selection and variability of evolutionary rates among mitochondrial lineages (Ballard & Kreitman, 1994; Rand *et al.*, 1994).

**Neutral expectations.** More generally, patterns of chloroplast genome partitioning and subsequent lineage sorting following speciation can explain the plastome polymorphism within populations of *dubia* and *georgiana*. Theoretical (Tajima, 1983; Nei, 1986) and numerical (Avice *et al.*, 1984; Neigel & Avice, 1986) studies have shown that for animal mtDNA, polyphyly and paraphyly are not unexpected patterns in recently diverged taxa. Polyphyletic patterns, in which a genome from one population is more closely related to a genome from a distant population than it is to a genome of a neighbouring individual, do not necessarily imply admixture or migration (Tajima, 1983; Nei, 1986; Pamilo & Nei, 1988). If an ancestral stock was polymorphic, a natural progression is expected from polyphyly through paraphyly to monophyly (Tajima, 1983; Avice *et al.*, 1987). A progression similar to this was observed in *Peromyscus* (Avice *et al.*, 1983) and Hawaiian *Drosophila* (DeSalle *et al.*, 1986).

Historical forces and selection appear to be the two most likely causes of the current cpDNA polymorphisms in populations of *Phacelia dubia*. In either case, evidence against either migration or recent admixture suggests that polymorphism predates taxon divergence. Fragmentation of a once more widespread and continuous habitat may explain the current distribution of identical haplotypes within different populations. Larger sample

sizes, including data on spatial structuring within populations and data on the relative fitnesses of individuals with different genomes under different conditions would help distinguish between an actively maintained polymorphism and one whose frequencies are governed primarily by drift.

#### *Phylogenetic implications*

**Ancestry of putative varieties.** The chloroplast genomes of both incipient varieties ('imitator' and 'railroad') of *Phacelia dubia* were clearly derived from a D plastome. The common cpDNA genotype in 'railroad', found in all but one individual sampled, occurred as part of a polymorphism in a nearby *dubia* population (d-2). The two 'railroad' populations have probably arisen from a *dubia* ancestor carrying the D plastome. In contrast, the two populations of 'imitator' may differ in their evolutionary histories. Despite their relationship to individuals with D plastomes, the two populations did not share a cpDNA haplotype. However, the predominant haplotype in each population was also found in a North Carolina population of *dubia* (population d-1). The common haplotype in 'imitator' population p-1 was also the most common haplotype in the entire survey, occurring in two *dubia* and three *georgiana* populations (Table 4) but the haplotype fixed in population p-2 was only found in a single *dubia* population (population d-1).

**Evolutionary relationships of recognized taxa.** Although there have been relatively few detailed studies of intraspecific diversity in cpDNA, the available data suggest that population genetic characteristics of the chloroplast genome are not unlike other genomes. That is, not surprisingly, mutations probably arise randomly within populations, they spread via migration, and drift and/or selection govern frequencies within populations. Furthermore, even with a limited number of studies, intraspecific surveys clearly indicate that different species display a wide spectrum of absolute levels of variation and these species vary in the relative partitioning of variation within and between populations. However trivial these conclusions may appear, they do highlight the need for intraspecific analyses prior to analyses at higher taxonomic levels. In this regard, cpDNA is not unlike any other taxonomic character.

The phylogeny of plastomes illustrated a clear evolutionary relationship in which plastomes D and I represent a distinct clade from plastomes G and M (Fig. 3a). However, caution must be exercised prior to equating the plastome and taxon phylogenies.

Intraspecific variations, as exemplified by the D–G plastome polymorphism, demonstrated two serious problems in taxonomic phylogeny reconstruction based solely upon cpDNA. Inadequate sampling could lead to an erroneous estimation of taxonomic relationships and, second, a phylogeny of genes is not necessarily concordant with a phylogeny of taxonomic entities. The latter can be especially troublesome when a phylogeny is derived from cpDNA because the chloroplast genome is essentially asexual and haplotypes are therefore best treated as alleles of a single locus (Doyle, 1992). If population surveys identify new haplotypes at a rate of one per eight to 25 individuals, then sample sizes should reflect these expected levels of polymorphism. Of particular relevance to the situation in *P. dubia* is the fact that discordant phylogenies are most likely to be observed when the taxa are recently derived. Plastome polymorphisms within populations of *dubia* and *georgiana* in addition to individuals with identical cpDNA haplotypes occurring in each of these two taxa suggest that the split in plastomes D and G probably predated divergence of the taxonomic varieties *dubia* and *georgiana*.

### Acknowledgements

We thank B. Anderson, B. Burkhardt, L. Harris, J. P. Hosler, P. Liu, K. MacElwaine and D. Robertson for helpful advice and technical assistance. B. Weir kindly supplied a computer program to estimate coancestry. The manuscript benefited from M. Uyenoyama's insightful comments and the suggestions of two anonymous reviewers. This research was supported in part by NIH grant no. 5 T32 GM07754-09, NSF grant no. 333-0861 and a Grant-in-aid-of-Research from Sigma Xi.

### References

- AVISE, J. C., ARNOLD, J. BALL, R. M., BIRMINGHAM, E., LAMB, T., NEIGEL, J. E., REEB, C. A. AND SAUNDERS, N. C. 1987. Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Ann. Rev. Ecol. Syst.*, **18**, 489–522.
- AVISE, J. C., NEIGEL, J. E. AND ARNOLD, J. 1984. Demographic influences on mitochondrial DNA lineage survivorship in animal populations. *J. Mol. Evol.*, **20**, 99–105.
- AVISE, J. C., SHAPIRA, J. F., DANIEL, S. W., AQUADRO, C. F. AND LANSMAN, R. A. 1983. Mitochondrial DNA differentiation during the speciation process in *Peromyscus*. *Mol. Biol. Evol.*, **1**, 38–56.
- BALLARD, J. W. O. AND KREITMAN, M. 1994. Unraveling selection in the mitochondrial genome of *Drosophila*. *Genetics*, **138**, 757–772.
- BANKS, J. A. AND BIRKY, C. W., JR. 1985. Chloroplast DNA diversity is low in a wild plant, *Lupinus texensis*. *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 6950–6954.
- BERMINGHAM, E. AND AVISE, J. C. 1986. Molecular zoogeography of freshwater fishes in the southeastern United States. *Genetics*, **113**, 939–965.
- BYRNE, M. AND MORAN, G. F. 1994. Population divergence in the chloroplast genome of *Eucalyptus nitens*. *Heredity*, **73**, 18–28.
- DESALLE, R., GIDDINGS, L. V. AND KANESHIRO, L. Y. 1986. Mitochondrial DNA variability in natural populations of Hawaiian *Drosophila*. II. Genetic and phylogenetic relationships of natural populations of *D. silvestris* and *D. heteroneura*. *Heredity*, **56**, 87–92.
- DOYLE, J. J. 1992. Gene trees and species trees: Molecular systematics as one-character taxonomy. *Syst. Bot.*, **17**, 144–163.
- ENNOS, R. A. 1994. Estimating the relative rates of pollen and seed migration among plant populations. *Heredity*, **74**, 250–259.
- GILLET, G. W. 1964. Genetic barriers in the *Cosmanthus Phacelias*. *Rhodora*, **66**, 359–366.
- HAMRICK, J. L., LINHART, Y. B. AND MITTON, J. B. 1979. Relationships between life history characteristics and electrophoretically detectable genetic variation in plants. *Ann. Rev. Ecol. Syst.*, **10**, 173–200.
- HARRIS, S. A. AND INGRAM, R. 1991. Chloroplast DNA and biosystematics: the effects of intraspecific diversity and plastid transmission. *Taxon*, **40**, 393–412.
- HONG, Y.-P., HIPKINS, V. D. AND STRAUSS, S. H. 1993. Chloroplast DNA diversity among trees, populations and species in the California closed-cone pines (*Pinus radiata*, *Pinus muricata* and *Pinus attenuata*). *Genetics*, **135**, 1187–1196.
- HOUGLANDER, N., LUMARET, R. AND BOS, M. 1993. Inter-intraspecific variation of chloroplast DNA of European *Plantago* spp. *Heredity*, **70**, 322–334.
- KEMBLE, R. J. 1987. A rapid, single leaf, nucleic acid assay for determining the cytoplasmic organelle complement of rapeseed and related *Brassica* species. *Theor. Appl. Genet.*, **73**, 364–370.
- KOCHERT, G., HALWARD, T., BRANCH, W. D. AND SIMPSON, C. E. 1991. RFLP variability in peanut (*Arachis hypogaea* L.) cultivars and wild species. *Theor. Appl. Genet.*, **81**, 565–570.
- LEVY, F. 1991a. Morphological differentiation in *Phacelia dubia* and *P. maculata*. *Rhodora*, **93**, 11–25.
- LEVY, F. 1991b. A genetic analysis of reproductive barriers in *Phacelia dubia*. *Heredity*, **67**, 331–345.
- MASON, R. J., HOLSINGER, K. E. AND JANSEN, R. K. 1994. Biparental inheritance of the chloroplast genome in *Coreopsis* (Asteraceae). *J. Hered.*, **85**, 171–173.
- MAYER, M. S. AND SOLTIS, P. S. 1994. The evolution of serpentine endemics: A chloroplast DNA phylogeny of the *Streptanthus glandulosus* complex (Cruciferae). *Syst. Bot.*, **19**, 557–574.
- MCCAULEY, D. E. 1994. Contrasting the distribution of chloroplast DNA and allozyme polymorphism among local populations of *Silene alba*: Implications for studies

- of gene flow in plants. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 8127–8131.
- MILLIGAN, B. G. 1991. Chloroplast DNA diversity within and among populations of *Trifolium pratense*. *Curr. Genet.*, **19**, 411–416.
- MURDY, W. H. 1966. The systematics of *Phacelia maculata* and *Phacelia dubia* var. *georgiana*, both endemic to granite outcrop communities. *Am. J. Bot.*, **53**, 1028–1036.
- NEALE, D. B., SAGHAI-MAROOF, M. A., ALLARD, R. W., ZHANG, Q. AND JORGENSEN, R. A. 1988. Chloroplast DNA diversity in populations of wild and cultivated barley. *Genetics*, **120**, 1105–1110.
- NEI, M. 1986. Stochastic errors in DNA evolution and molecular phylogeny. In: Gershowitz, H., Rucknagel, D. L. and Tashian, R. E. (eds) *Evolutionary Perspectives and the New Genetics*, pp. 133–147. Alan Liss, New York.
- NEIGEL, J. E. AND AVISE, J. C. 1986. Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. In: Nevo, E. and Karlin, S. (eds) *Evolutionary Processes and Theory*, pp. 515–534. Academic Press, New York.
- PALMER, J. D. 1982. Physical and gene mapping of chloroplast DNA from *Atriplex triangularis* and *Cucumis sativa*. *Nucl. Acids Res.*, **10**, 1593–1605.
- PALMER, J. D. 1985. Comparative arrangement of chloroplast genomes. *Ann. Rev. Genet.*, **19**, 325–354.
- PALMER, J. D. 1987. Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. *Am. Nat.*, **130**, S6–S29.
- PALMER, J. D. AND ZAMIR, D. 1982. Chloroplast DNA evolution and phylogenetic relationships in *Lycopersicon*. *Proc. Natl. Acad. Sci. U.S.A.*, **79**, 5006–5010.
- PALMER, J. D., JANSEN, R. K., MICHAELS, H. J., CHASE, M. W. AND MANHART, J. R. 1988. Chloroplast DNA variation and plant phylogeny. *Ann. Mo. Bot. Gard.*, **75**, 1180–1206.
- PAMILO, P. AND NEI, M. 1988. Relationships between gene trees and species trees. *Mol. Biol. Evol.*, **5**, 568–583.
- PETIT, R. J., KREMER, A. AND WAGNER, D. B. 1994. Geographic structure of chloroplast DNA polymorphisms in European oaks. *Theor. Appl. Genet.*, **87**, 122–128.
- RAND, D. M., DORFSMAN, M. AND KANN, L. M. 1994. Neutral and non-neutral evolution of *Drosophila* mitochondrial DNA. *Genetics*, **138**, 741–756.
- SCOWCRAFT, W. R. 1979. Nucleotide polymorphism in chloroplast DNA of *Nicotiana debneyi*. *Theor. Appl. Genet.*, **55**, 133–137.
- SLATKIN, M. 1985. Rare alleles as indicators of gene flow. *Evolution*, **39**, 53–65.
- SOLTIS, D. E., SOLTIS, P. S., KUZOFF, R. K. AND TUCKER, T. L. 1992. Geographic structuring of chloroplast DNA genotypes in *Tiarella trifoliata* (Saxifragaceae). *Pl. Syst. Evol.*, **181**, 203–216.
- SWOFFORD, D. L. 1991. PAUP: *Phylogenetic Analysis Using Parsimony*, version 3.0. Computer program distributed by the Illinois Natural History Survey.
- TAJIMA, F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437–460.
- TERAUCHI, R., TERAUCHI, T. AND TSUNEWAKI, K. 1991. Intraspecific variation of chloroplast DNA in *Dioscorea bulbifera* L. *Theor. Appl. Genet.*, **81**, 461–470.
- WEIR, B. S. 1990. *Genetic Data Analysis*. Sinauer, Sunderland, MA.
- WOLFE, K. H., LI, W.-H. AND SHARP, P. M. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. U.S.A.*, **84**, 9054–9058.
- WOLFF, K. AND SCHAAL, B. 1992. Chloroplast DNA variation within and among five *Plantago* species. *J. Evol. Biol.*, **5**, 325–344.