Theor Appl Genet (1999) 99:859–867

G.J. Bryan · J. McNicoll · G. Ramsay · R.C. Meyer
W.S. De Jong

Polymorphic simple sequence repeat markers in chloroplast genomes of Solanaceous plants

Received: 15 December 1998 / Accepted: 9 February 1999

Abstract PCR-based markers were developed from mononucleotide simple-sequence repeats in the chloroplast genome of *Nicotiana tabacum* and applied to the analysis of genetic diversity. These markers were found to detect high levels of polymorphism at three taxonomic levels in Solanaceous plants. Of 36 chloroplast loci examined, 26 show some degree of polymorphism among potato accessions. Among a set of 30 tetraploid potato cultivars it is apparent that a single chloroplast haplotype is prevalent, presumably a result of the widespread use as a female parent of the imported US cultivar Rough Purple Chili in the latter half of the 19th century. Nonetheless, there is considerable chloroplast diversity in the cultivated potato, and it is clear that a large proportion of this variability has arisen through the use of wild or primitive cultivated species of potato in introgression programmes. This variability should be used in future breeding programmes. An examination of single accessions from 24 potato species, as well as representatives from tobacco and other members of the Solanaceae, reveals high levels of inter-specific chloroplast DNA variation. These data, and the ease of use and potential for multiplexing of these markers, suggest that cpSSRs will be of great utility in population genetics, germplasm management, evolutionary and phylogenetic studies as well as in, the analysis of material from introgression and somatic-fusion experiments. Interestingly, the polymorphism arising from one of the more-polymorphic chloroplast loci examined, does not originate solely from the SSR, and is due to variation in the copy number of two tandemly arrayed sequence elements.

Key words Solanaceae · Microsatellite · Simple sequence repeat · Chloroplast genome · PCR-based markers

Communicated by J.W. Snape

G.J. Bryan (✉) · J. McNicoll · G.Ramsay · R.C. Meyer
W.S. De Jong
Crop Genetics Department, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK

Introduction

The use of information from organelle genomes is well established in the study of population and evolutionary processes in both animals and plants (Avise 1994). Genetic markers derived from organelle genomes generally show simple, uniparental modes of inheritance, which makes them invaluable for the purposes of population-genetic and phylogenetic studies. In plants, these studies have typically been based upon chloroplast genome-derived RFLPs or sequence data (Soltis et al. 1992). More recently, genetic markers based upon simple-sequence repeats in chloroplast genomes (cpSSRs) have been shown to be highly useful markers for the study of germplasm analysis in several plant species, such as soybean (Powell et al. 1996), rice (Provan et al. 1996), and pine (Powell et al. 1995). These SSRs are analogous to those in the nuclear genome, except that cpSSRs are characteristically composed of mononucleotide repeats rather than di-, tri- or tetra-nucleotide repeats. Owing to the haploid nature and high copynumber of the chloroplast genome, these markers are particularly easy to deploy and analyse using PCR and polyacrylamide gel electrophoresis.

Cultivated potato is the world's fourth most important food crop and there are seven cultivated, and over 200 wild species of potato. The major cultivated species is *Solanum tuberosum*, which consists of two subspecies *S. tuberosum* ssp. *tuberosum* and *S. tuberosum* ssp. *andigena*. Current knowledge of the phylogenetic relationships and origins of cultivated potato are based upon morphological data (Hawkes 1990) and limited amounts of molecular marker data, largely nuclear (Debener et al. 1990)- or chloroplast (Hosaka et al. 1984; Spooner et al. 1991)-derived RFLP markers. Previous studies have examined the levels of chloroplast restriction-site DNA variation in wild and cultivated species of potato (Hosaka and Hanneman 1988). *S. tuberosum* ssp. *tuberosum* shows low levels of chloroplast DNA variation; nevertheless; it was possible to subdivide potato cultivars into three groups (Powell et al. 1993). Other studies have used cpDNA to address questions concerning the mater-

nal origins of cultivated potato and the levels of chloroplast variation among both wild and cultivated species of potato (Hosaka 1986; Hosaka and Hanneman 1988a, b). Chloroplast restriction-site data have also been used to construct a phylogenetic tree of the Solanaceae (Olmstead and Palmer 1992).

In this study we present the development of 36 SSR loci from the chloroplast genome of *Nicotiana tabacum* and their application to potato and tobacco germplasm analysis. These markers reveal high levels of both intra- and inter-specific diversity and should find applications in the study of genetic diversity and population-genetic processes in a wide range of Solanaceous and other plant species.

Materials and methods

Plant material and DNA extractions

DNA was extracted from fresh or frozen plant tissue by a modified version of the CTAB method (Murray and Thompson 1980). Leaf tissue was taken from potato cultivars, breeding lines or from wild or primitive wild species from the Commonwealth Potato Collection.

The samples included in this study are as follows:

Potato (*S. tuberosum* ssp. *tuberosum*) cultivars: Brodick, Pentland Dell, Stirling, Désirée, Pentland Crown, Maris Piper, Record, Torridon, Brodie, Cara, Pentland Squire, Pentland Ivory, Bintje, King Edward, Russet Burbank, Perricholi, Lumpers, Pink Fir Apple, Shelagh, Yam, Centifolia, Peach Bloom, Kerrs Pink, Arran Consul, Majestic, Myatts Ashleaf, Skirza.

Potato (*S. tuberosum* ssp. *tuberosum*) breeding lines: P55/7, 62.33.3, 65.346.19

Wild and primitive cultivated species of potato (accession number): *Solanum sparsipilum* CPC3564 (spl), *Solanum phureja* IVP48 (phu), *Solanum acaule* CPC3735 (acl), *Solanum chacoense* CPC3057 (chc), *Solanum demissum* CPC7152 (dms), *Solanum pinnatisectum* CPC3234 (pnt), *Solanum papita* CPC7081 (pta), *Solanum stoloniferum* CPC7110 (sto), *Solanum stenotomum* CPC4703 (stn), *S. stenotomum* ssp. *goniocalyx* CPC4655 (stn gon), *Solanum sanctae rosae* CPC3779 (sct), *Solanum sandemanii* CPC6027 (snd), *Solanum venturii* CPC3715 (vnt), *Solanum kurtzianum* CPC4039 (ktz), *Solanum medians* CPC3274 (med), *S. tuberosum* ssp. *andigena* CPC303 (adg), *Solanum bukasovii* CPC2438 (buk), *Solanum mochiquense* CPC7182 (moc), *Solanum berthaultii* BPC106 (ast), *Solanum aemulans* CPC7012 (aem), *Solanum circaefolium* ssp. *quimense* CPC7088/7089 (qum), *Solanum verrucosum* CPC7213 (ver), *Solanum brevidens* CPC2451 (brd), *Solanum vernei* CPC4078 (vrn).

Other Solanaceous plant accessions: Tomato (*Lycopersicon esculentum* cv Moneymaker), cape gooseberry (*Physalis floridiana*), *Petunia* sp., pepper (*Capsicum* sp.), aubergine (*Solanum melongena*), *Nicotiana edwarsonii*, *Nicotiana clevelandii*, *Nicotiana benthami-*

Table 1 Primers raised to the *N. tabacum* chloroplast genome

Locus	Gene location of microsatellite repeat(s)	Repeat	Primers (5'–3')	Position in Nt cpDNA	Expected Size	T_m (°C)
NTCP2	*trn*H/*psb*A intergenic region	T_{10}·A_{10}	CTCGCCTACTTACATTCC AAGGAGAGGTTATTTTCTTG	83–100 376–357	294	50
NTCP3	*trn*K intron	T_{10}	AAGTCAAAAGAGCGATTAG TGATACATAGTGCGATACAG	3,903–3,884 3,707–3,727	196	50
NTCP4	*trn*K/*rps*16 intergenic region	A_{12}	TTGGATTAGATTTGTAGTTCCA ATCCACTTCATTTATCACAATG	4,692–4,671 4,536–4,557	162	55
NTCP5	*rps*16/*trn*Q intergenic region	T_{14}	CGAATTGATAGATACGAAACC AATACACCAAACAACAAATCC	6,861–6,881 6,732–6,752	150	54
NTCP6	*rps*16/*trn*Q intergenic region	A_{11}	GGTTCGAATCCTTCCGTC GATTCTTTCGCATCTCGATTC	7,437–7,420 7,262–7,282	176	60
NTCP7	ORF98/*trn*S intergenic region	A_{11}	TGATCCCGGACGTAATCC CGAATCCCTCTCTTTCCG	8,475–8,492 8,649–8,632	175	60
NTCP8	*trn*G intron	T_{11}	ATATTGTTTTAGCTCGGTGG TCATTCGGCTCCTTTATG	9,895–9,914 10,145–10,138	251	55
NTCP9	*trn*G/*trn*R intergenic region	T_{10}	CTTCCAAGCTAACGATGC CTGTCCTATCCATTAGACAATG	10,220–10,237 10,456–10,435	237	55
NTCP10	*atp*F intron	T_{13}	TGCTGAATCGACGACCTA AATATTCGGAGGACTCTTCTG	12,931–12,914 12,812–12,832	120	55
NTCP11	*ars*2	T_{10}	AGTGAATATTCATTGAGACGAACG ATCTAGAGTGATAGCAAAAA	15,007–15,028 15,142–15,121	136	60
NTCP12	*rps*2/RF862 intergenic region	T_{10}·A_{13}	CCTCCATCATCTCTTCCAA ATTTATTTCAGTTCAGGGTTCC	16,893–16,911 17,018–16,997	126	60
NTCP13	*rpo*C2 exon	T_{10}	TTTCCTGTTCCTGGTGGTA TTGGGGTAGATACACAAATCAC	19,095–19,113 19,272–19,251	178	60
NTCP14	*psb*M/*trn*D intergenic region	T_{11}	AATCCGTAGCCAGAAAAATAAA CCGATGCATGTAATGGAATC	31,571–31,592 31,722–31,703	152	60
NTCP15	*trn*E/*trn*T intergenic region	A_{10}	TGCAAAAATCCTACTCTTC TCTTTCGATTATTTTAGTTTAGA	32,654–32,636 32,535–32,557	120	50
NTCP16	*trn*E/*trn*T intergenic region	T_{13}	TCTAAACTAAAATAATCGAAAGA TGAAATTGTCAATATAATCGA	32,557–32,535 32,373–32,393	185	50
NTCP18	*psb*C/*trn*S intergenic region	T_{11}	CTGTTCTTTCCATGACCCCTC CCACCTAGCCAAGCCAGA	36,864–36,884 37,049–37,032	186	60
NTCP19	*ycf*3 intron	T_{17}	AATCGTTGTTTTAGACGATGC GAAACCCATTCTTACCACAAG	45,055–45,075 45,219–45,199	165	55
NTCP20	*ycf*3 intron	A_{13}	TCCTCGTAAGACTGAGAGAAAT TTACGAGTAATTCCGACAACTT	46,071–46,092 46,192–46,171	122	60

Table 1 Continue

Locus	Gene location of microsatellite repeat(s)	Repeat	Primers (5'–3')	Position in Nt cpDNA	Expected Size	T_m (°C)
NTCP21	ORF74A exon	T_{10}	AAAAAGATCCCACAAAGAAAA CTTATCGATTCCTGTCAAAAAG	46,303–46,323 46,430–46,409	128	60
NTCP22	ORF74A/*trn*S intergenic region	T_{10}	TATCAGAAAAAGAAAAAGAAGG GTCAAAGCAAAGAACGATT	46,868–46,889 46,999–46,981	132	60
NTCP23	*rps*4/*trn*T intergenic region	A_{10}	CAAAGGAACATTATCAATCATC TAGCTCAGAGGTTAGAGCATC	48,449–48,470 48,570–48,551	122	55
NTCP24	*atp*B exon	T_{10}	GACCGATGATTTGGACGAC GCTAGCGGACATTTATTTTGAA	56,685–56,703 56,841–56,820	157	60
NTCP25	*atp*B/*rbc*L intergenic region	A_{13}	TTAGTCAGGTATTTCCATTTC CTTTTCATAGGAATCTTTCACA	57,318–57,338 57,507–57,486	190	50
NTCP26	*psa*I/ORF184 intergenic region	T_{10}	GCAATTGCAATGGCTTCTTTA TTTATGTTCGGTGGAAATCACA	62,132–62,152 62,299–62,278	168	60
NTCP27	*trn*P/*psa*J intergenic region	T_{10}	ATAAATACAGAACCCGTCGTAA TGCTTAGAGTTGGACACAGAAT	69,314–69,335 69,479–69,458	166	60
NTCP28	*rpl*20/*rps*12 intergenic region	T_{14}	TCCAATGGCTTTGGCTA AGAAACGAAGGAACCCAC	71,518–71,534 71,687–71,670	170	57
NTCP29	*clp*P intron	A_{10}	AGTCGGTTGATTAGGGTAAAAT AAAGCCCTTTCGTTAGAAGTAA	73,618–73,639 73,774–73,753	157	60
NTCP30	*clp*P intron	T_{13}·T_{15}	GATGGCTCCGTTGCTTTAT TGCCGGAGAGTTCTTAACAATA	72,941–72,959 73,103–73,082	158	60
NTCP32	*clp*P intron	T_{11}	TGTTTCATCTTTTTAGGTTTAT TCAAGCAAAGTTATCTCAAC	73,955–73,976 74,119–74,100	165	60
NTCP33	*rpo*A exon	T_{10}	TGGCTGTTATTCAAAAGGTC CATGATAAATTGGCTAAACTCA	80,558–80,577 80,706–80,685	149	60
NTCP34	*rps*8/*rp*l14 intergenic region	T_{10}	GATCCATATCAGCATTTCGTAT CTGTTTCTAGTGGGGTATTTGA	82,947–82,968 83,158–83,137	212	60
NTCP36	*rps*19/*rp*l2 intergenic region	T_{14}	GTAGTAAATAGGAGAGAAAATAGA TGATACATAGTGCGATACAG	86,751–86,728 86,634–86,654	125	50
NTCP37	*rrn*5/*trn*R intergenic region	A_{13}	TTCCGAGGTGTGAAGTGG CAGGATGATAAAAAGCTTAACAC	109,852–109,835 109,710–109,732	143	55
NTCP38	*ycf*1/*ori*R intergenic region	T_{10}	AATTTCGAGGTTCTTATTTACT GATTCTGAGCTCTATTCATTAG	127,295–127,316 127,447–127,426	153	60
NTCP39	*trn*R/*rrn*5 intergenic region	T_{13}	GTCACAATTGGGGTTTTGAATA GACGATACTGTAGGGGAGGTC	132,707–132,728 132,862–132,842	156	60
NTCP40	*rp*12/*trn*H intergenic region	A_{14}	TAATTTGATTCTTCGTCGC GATGTAGCCAAGTGGATCA	155,758–155,776 76–58	163	55

ana, Nicotiana glutinosa, Nicotiana debneyii, Nicotiana occidentalis, Nicotiana tabacum cvs Xanthi, Samsun, White Burley.

PCR primers raised against a Nicotiana cpDNA sequence

The *N. tabacum* (EMBL accession CHNTXX) chloroplast genome (155 844-bp) sequence contains 39 poly(A/T)$_n$ SSRs of at least 10 bp in length. The GCG program FINDPATTERNS was used to locate these repeats in the chloroplast-genome sequence. This sequence was edited to form short (circa 500-bp) sequence segments containing the SSR arrays. The program OLIGO 5.0 (National Biosciences Inc., Plymouth, Minn. USA) was used to design oligonucleotides which flanked SSRs. Primers were designed to allow the generation of PCR products 100–300-bp in length, and PCR annealing temperatures of 50–60°C.

Polymerase chain reaction amplification

Genomic sequences were routinely amplified in 20-µl PCR reactions containing 25 ng of template DNA, in the presence of 20 mM Tris-HCl pH 8.4, 1.5 mM $MgCl_2$, 50 mM KCl, 0.05% (v/v) W1, 200 nM of each primer, 100 µM of each dNTP, and 0.4 units of *Taq* DNA polymerase. Cycling conditions were as follows: initial denaturation step of 4 min at 94°C, followed by 30 cycles of 30 s denaturation at 94°C, 1-min annealing at the appropriate T_m, and 1-min extension at 72°C. After cycling, PCR reactions were incubated for 5 min at 72°C. PCR products were analysed by electrophoresis on 1×TAE agarose gels and visualised by ethidium bromide staining.

For the analysis of polymorphism PCR products were subjected to electrophoresis on sequencing gels [6% polyacrylamide (19:1 acrylamide:bis), 8 M urea] run under standard conditions. Products were visualised either by silver staining (Promega Inc.), or by use of γ^{32}P-labelled PCR primers and exposure of dried gels to X-ray film. For accurate size determination, PCR products were run against a sequencing ladder.

DNA sequencing

PCR products were ligated onto pGEM-T Easy (Promega Inc.) and transformed into *Escherichia coli* strains DH5α or XL1-Blue by electroporation. DNA sequencing was performed using a Perkin Elmer Dye Terminator sequencing kit and analysed on a Perkin Elmer model 373A DNA sequencer.

Fragment scoring and statistical analysis

Statistical analysis was performed using the NTSYS-pc set of programs (Applied Biostatistics Inc.) or GENSTAT 5 Release 3.22. PCR products were scored, where possible, as the absolute of DNA fragment size. The informativeness of each PCR primer pair was calculated using the method of Nei (1986):

$$\hat{H} = \frac{n}{n-1}(1 - \sum p_i^2),$$

where p_i=frequency of ith allele, n=no. of samples.

Genetic similarity was calculated for each pair of accessions using the "city block" similarity coefficient. Cluster analysis was performed using the UPGMA method of Sneath and Sokal (1973).

Results

Levels of polymorphism shown by chloroplast SSRs

Owing to the very close proximity of three pairs of the 39 SSRs in the *N. tabacum* chloroplast genome, a total of 36 primer pairs were constructed. Table 1 lists the primer sequences and their locations in the *N. tabacum* chloroplast genome. Primers were assayed for polymorphism on a sample of 30 potato cultivars and breeding lines, and 24 wild and primitive cultivated potato species. The potato cultivars and breeding lines were selected to adequately represent the three cytoplasmic groups seen in an earlier study using chloroplast RFLPs to examine potato cytoplasmic diversity (Powell et al. 1993). The wild and primitive cultivated material was chosen to represent single accessions from a range of species within the Commonwealth Potato Collection.

Four of the 36 primer pairs (ntcp5, ntcp15, ntcp16, ntcp38) did not produce amplification products of the expected size from potato, although in all cases these primers produced a fragment of the predicted size from tobacco (data not shown). Six primer pairs (ntcp2, ntcp13, ntcp19, ntcp24, ntcp33, ntcp34) failed to detect polymorphism among the set of 54 potato accessions. The remaining 26 primer pairs detected some level of polymorphism among the set of 54 potato accessions. Information on the levels of polymorphism among potato species and cultivars for the 32 loci for which PCR products could be generated is provided in Table 2. Among the set of potato germplasm the number of distinct alleles at each polymorphic locus ranged from 2 to 8 (mean=3.8) and heterozygosity values from 0.07 to 0.7 (mean=0.38).

All 17 primer combinations tested on tobacco and other Solanaceous plants, including tomato, pepper and aubergine, detected PCR products of the approximate expected size, and 3–9 alleles from a sample of 12 diverse accessions. Generally higher levels of polymorphism are seen among the nine tobacco accessions tested in this study, when compared with the results of the analysis based upon primitive cultivated and wild potato species (average heterozygosity value of 0.79 for tobacco versus 0.47 for wild potato species for same set of 17 primer pairs).

Figure 1 shows examples of the length variability in PCR fragments produced by cpSSR primers on potato germplasm (1a and 1b), as well as from tobacco and representatives of other Solanaceous plant species (1c).

Cluster analysis of chloroplast SSR haplotypes

The data from all of the SSR primers were scored for absolute fragment size for the sets of potato and tobacco germplasm described. Similarity matrices, constructed by the "city block" coefficient, were used to cluster accessions according to combined SSR locus haplotypes. The potato germplasm surveyed comprised 43 distinct chloroplast haplotypes. Among the *S. tuberosum* ssp. *tuberosum* cultivars and breeding material there were 19 distinct haplotypes from 30 accessions, whereas in the wild and primitive cultivated species there were 24 distinct haplotypes, each sample possessing a unique haplotype. The most striking feature of a cluster analysis of the *S. tuberosum* ssp. *tuberosum* material is a group of 12 cultivars, with identical or near-identical haplotypes (Fig. 2a). This cytoplasmic type clearly corresponds to the T-type cytoplasm described by Hosaka and Hanneman (1988b), as this analysis contains several cultivars included in an earlier study of tetraploid potato cultivars (Powell et al. 1993). Also evident is a second group of 11 cultivars with similar haplotypes to the cluster possessing identical haplotypes. Another cluster, possessing more diverse haplotypes, contains some older cultivars (eg. Lumpers, Pink Fir Apple) as well as a breeding line,

Table 2 Levels of polymorphism detected at cpSSR loci in potato germplasm

Locus	Av. het (no. of alleles)			Locus	Av. het (no. of alleles)		
	Cultivars	Wild	Overall		Cultivars	Wild	Overall
NTCP2	0.00 (1)	0.00 (1)	0.00 (1)	NTCP22	0.07 (2)	0.08 (2)	0.07 (2)
NTCP3	0.20 (3)	0.66 (4)	0.59 (4)	NTCP23	0.49 (3)	0.74 (5)	0.66 (5)
NTCP4	0.41 (3)	0.42 (3)	0.62 (4)	NTCP24	0.00 (1)	0.00 (1)	0.00 (1)
NTCP6	0.43 (4)	0.72 (6)	0.59 (7)	NTCP25	0.07 (2)	0.40 (2)	0.46 (2)
NTCP7	0.53 (3)	0.57 (3)	0.58 (3)	NTCP26	0.07 (2)	0.08 (2)	0.07 (2)
NTCP8	0.49 (4)	0.79 (5)	0.66 (5)	NTCP27	0.31 (3)	0.45 (3)	0.37 (3)
NTCP9	0.40 (4)	0.82 (8)	0.70 (8)	NTCP28	0.00 (1)	0.41 (3)	0.20 (3)
NTCP10	0.07 (2)	0.23 (2)	0.17 (2)	NTCP29	0.07 (2)	0.16 (2)	0.11 (2)
NTCP11	0.07 (2)	0.16 (2)	0.14 (3)	NTCP30	0.38 (2)	0.51 (4)	0.54 (4)
NTCP12	0.23 (2)	0.76 (5)	0.58 (5)	NTCP32	0.07 (2)	0.24 (3)	0.15 (3)
NTCP13	0.00 (1)	0.00 (1)	0.00 (1)	NTCP33	0.00 (1)	0.00 (1)	0.00 (1)
NTCP14	0.46 (3)	0.74 (5)	0.69 (5)	NTCP34	0.00 (1)	0.00 (1)	0.00 (1)
NTCP18	0.54 (3)	0.64 (3)	0.66 (3)	NTCP36	0.00 (1)	0.16 (2)	0.07 (2)
NTCP19	0.00 (1)	0.00 (1)	0.00 (1)	NTCP37	0.07 (2)	0.30 (3)	0.17 (3)
NTCP20	0.00 (1)	0.16 (2)	0.07 (2)	NTCP39	0.48 (3)	0.43 (4)	0.62 (4)
NTCP21	0.07 (2)	0.08 (2)	0.07 (3)	NTCP40	0.00 (1)	0.37 (3)	0.18 (3)

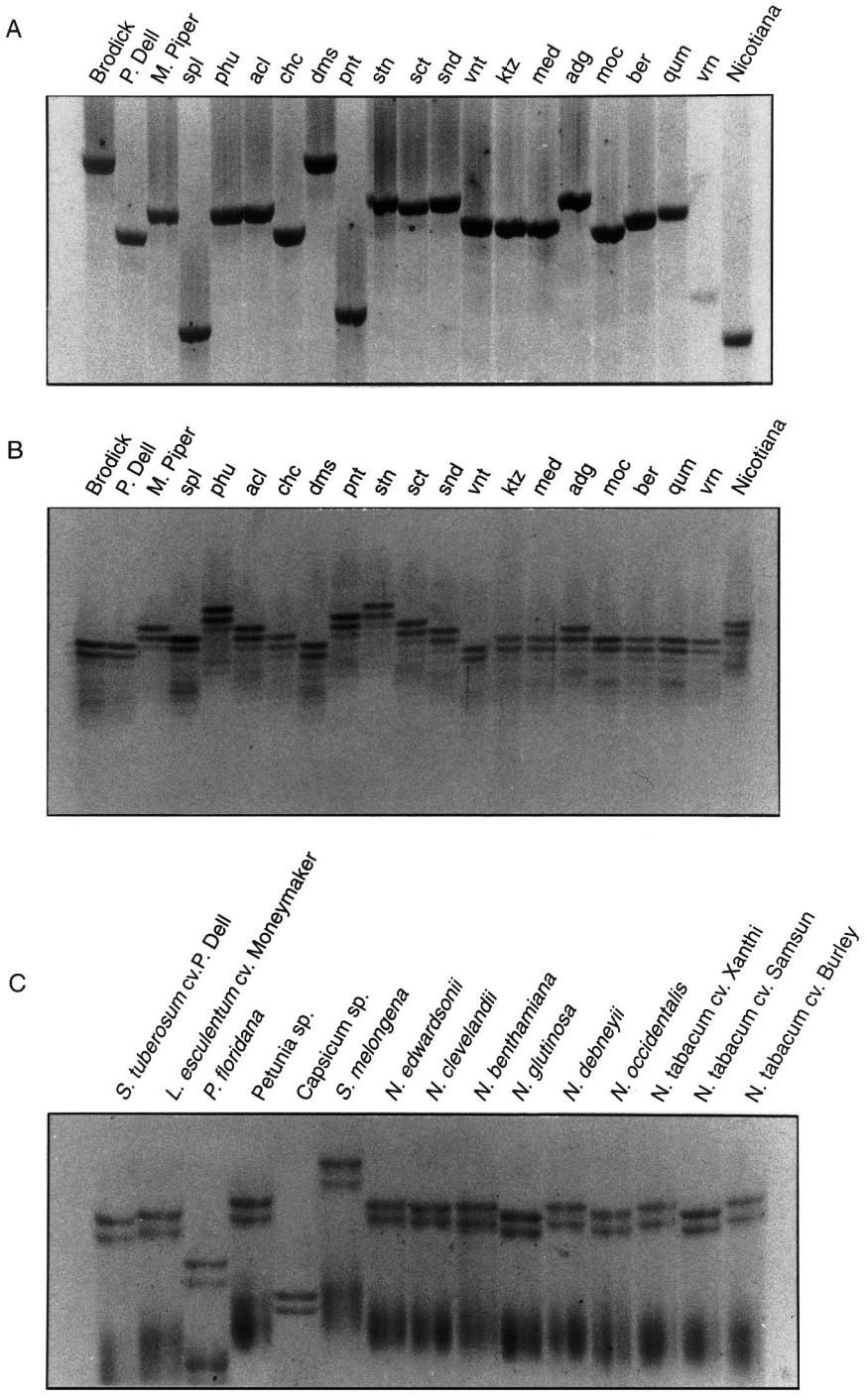


Fig. 1A–C Silver-stained sequencing gels showing polymorphic PCR products obtained using chloroplast SSR primers: **A** potato germplasm evaluated using primer pair NTCP9, **B** potato germplasm evaluated using primer pair NTCP12, and **C** Solanaceous plant species germplasm evaluated using primer pair NTCP12

P55/7. A third cluster contains two cultivars (Brodick and Shelagh) as well as the breeding line 65/346/19. A recent study of 179 cultivars on the UK National List has shown that the T-type cytoplasm predominates, with 151 (84%) possessing an identical chloroplast haplotype, based on a set of seven of the most polymorphic SSR markers presented here (Provan et al. 1998). The relatively high level of diversity seen among cultivars in this study reflects the deliberate selection of more-diverse chloroplast haplotypes from a previous analysis of chloroplast diversity in the cultivated potato (Powell et al. 1993). It is particularly noticeable that the level of cpSSR polymorphism seen within the cultivated *S. tuberosum* ssp. *tuberosum* gene pool is considerably higher than that seen in previous studies using other types of chloroplast marker (e.g. Powell et al. 1993).

Similarity coefficients for the accessions from 24 wild and primitive species reveal that each accession has a unique haplotype (data not shown). The clustering patterns are in good agreement with results obtained from phylogenetic studies of potato. A major cluster containing species of Mexican origin is apparent. As expected,

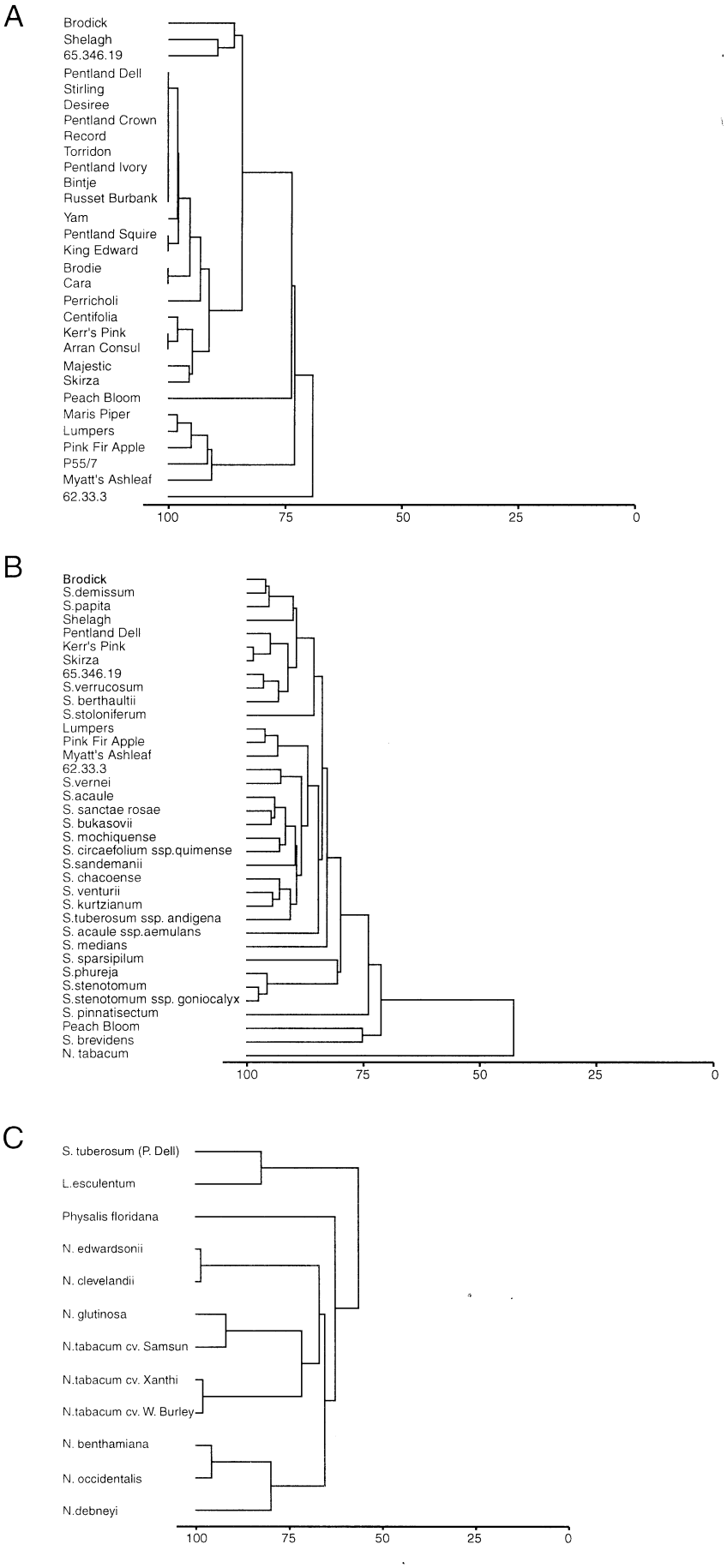
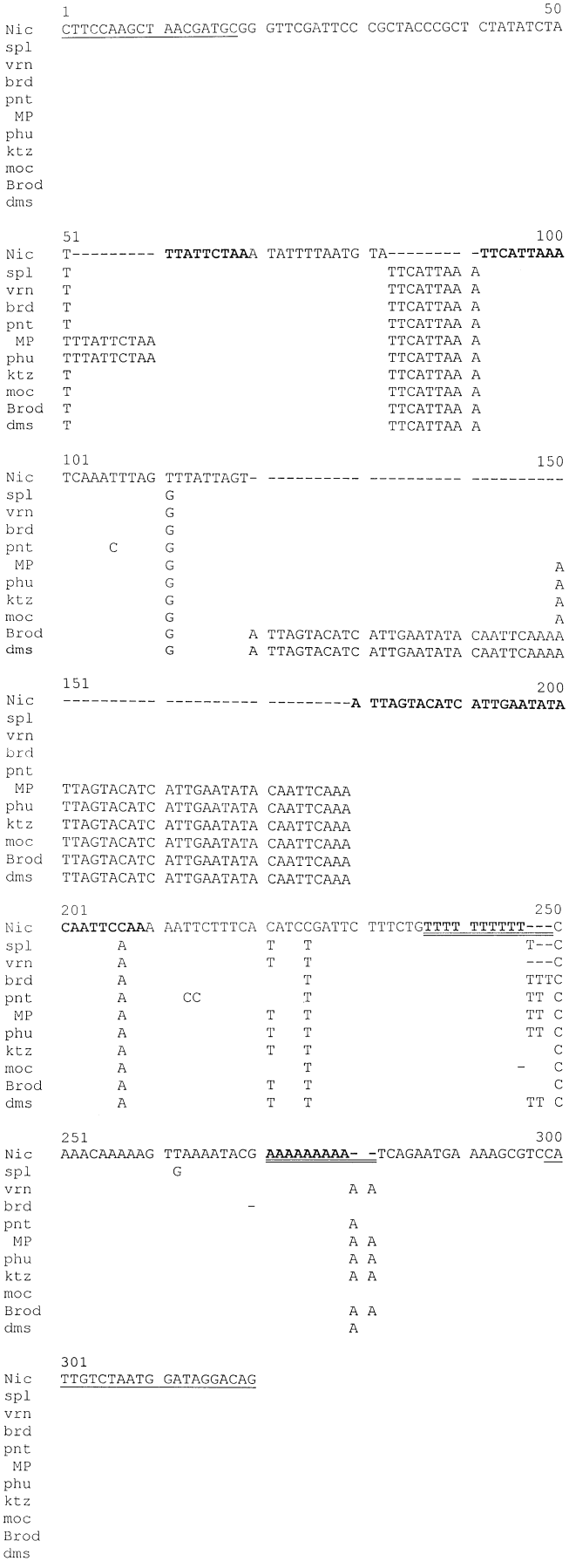


Fig. 2A–C Dendrograms showing UPGMA clustering of 30 potato cultivar haplotypes (**A**), distinct potato cultivar haplotypes and wild/primitive cultivated species haplotypes (**B**) and Solanaceous plant species haplotypes, including cultivars of *N. tabacum* (**C**)



the closely related species *S. phureja*, *S. stenotomum* and *S. stenotomum* ssp. *goniocalyx* cluster together. On merging the combined haplotypes from wild and primitive cultivated species with those from *S. tuberosum* ssp. *tuberosum* haplotypes representing the major groupings shown in Fig. 2a, it is possible to determine to what extent wild-species chloroplast genomes have been introgressed into the cultivated potato. It is clear that several of the potato cultivars possess cpDNA haplotypes which cluster more closely with haplotypes of particular wild or primitive cultivated species than with those other potato cultivars (Fig. 2b). Breeding records indicate that the cultivar Brodick has a pedigree containing an accession from the species *S. demissum*. The close correspondence of the cpDNA haplotypes of Brodick and Shelagh to those of *S. demissum* and *S. papita* accessions would suggest that one of these species, or at least a closely related Mexican species, was used as a female parent in the breeding history of these cultivars. Similarly, the breeding line 62.33.3 is known to contain genetic material introgressed from the wild species *S. vernei* and it is apparent that their chloroplast haplotypes are very similar. Another breeding line, 65.346.19, possesses a haplotype similar to the accession of *S. verrucosum* used in this study. A small set of potato cultivars (e.g. Maris Piper, Lumpers, Pink Fir Apple), previously described to possess a chloroplast DNA type derived from *S. tuberosum* ssp. *andigena*, cluster more closely with a large proportion of the wild and primitive cultivated species examined in this study, including *S. tuberosum* ssp. *andigena*.

Cluster analysis based on 17 cpSSRs for accessions from 12 Solanaceous plant species, comprising single accessions of potato, tomato and cape gooseberry, and nine Nicotiana sp. accessions, was also performed. The close similarity between the chloroplast genomes of *N. edwardsonii* and *N. clevelandii* suggests a common cytoplasmic origin. A similar situation exists for the tobacco species *N. benthamiana* and *N. occidentalis*. The N gene for TMV was introgressed into *N. tabacum* from *N. glutinosa*. The cultivar *N. tabacum* cv Samsun, which carries the N gene, possesses a cytoplasmic haplotype most similar to that of *N. glutinosa*.

Non-microsatellite-derived polymorphism at chloroplast SSR loci

Most of the cpSSR primers detect the typical “continuous” pattern of length variability, showing “stuttering” and small (i.e. 1—2 bp) differences in molecular weight

Fig. 3 Sequence alignment of cloned NTCP9 PCR products using PILEUP. Genotypes are as follows: *Nic=Nicotiana tabacum*; *spl=S. sparsipilum*; *vrn=S. vernei*; *brd=S. brevidens*; *pnt=S. pinnatisectum*; *MP=Solanum tuberosum* ssp. *tuberosum* cv Maris Piper; *phu=S. phureja*; *ktz=S. kurtzianum*; *moc=S. mochiquense*; *Brod=Solanum tuberosum* ssp. *tuberosum* cv Brodick; *dms=S. demissum*. Sequences *singly underlined* denote PCR primers. Simple sequence repeats are *double-underlined*. Sequences present in one copy in *Nicotiana* but 1–3 times in accessions of Solanum are in *bold face* type

suggesting that the polymorphism is a direct reflection of length variation in the mononucleotide simple-sequence repeat. Sequencing of PCR products from cpSSR loci has generally confirmed this supposition (data not shown). However, for a small number of chloroplast loci, the relative sizes of the PCR fragments suggested that insertions and deletions of small segments of DNA, not directly associated with the SSR, were responsible for the observed length variability. For example, the most polymorphic locus NTCP9 gave a fragment pattern which suggests that the very high level of polymorphism is not entirely due to variation in the length of the mononucleotide repeat. Subsequent sequencing of PCR products revealed that this locus contains tandem arrays of 9 and 30 bp in length which also vary in copy number at this locus and which are responsible for the "discontinuous" array of allele sizes. In the *Nicotiana* chloroplast genome this fragment contains only a single copy of each of these repeats, whereas potato accessions contain 1–2 (9-bp) or 1–3 (30 bp) copies of these sequences. Figure 3 shows alignments of the sequences of some of the PCR products generated by the NTCP9 primers. This suggests that this type of larger insertion can "mask" polymorphism arising from the microsatellite present at this locus. It is also evident that a second SSR is showing polymorphism at NTCP9. This second SSR is 9-bp in length in the sequenced *Nicotiana* chloroplast genome, and is 9–12-bp in length in the potato accessions examined at the sequence level. A second locus (NTCP40) generates this type of discontinuous variation among *Nicotiana* accessions but shows the more-typical pattern of variation among potato accessions (data not shown).

Discussion

In this study we have shown that PCR primers raised to *Nicotiana* cpDNA SSRs are able to detect length variation within the chloroplast genomes of tobacco, potato and other Solanaceous plants. Our preliminary data suggest that the high levels of intra- and inter-specific diversity seen in potato is exceeded if material from other Solanaceous plant genera is subjected to a similar type of analysis. Owing to the high levels of polymorphism and ease of applicability, these markers are likely to be of particular utility in the study of population-genetic processes in wild populations of Solanaceous plants and for the rapid analysis of germplasm collections containing several species. Our data show that 26 of 36 pairs of primers detect polymorphism among potato accessions. It is probable that a greater proportion of these 36 loci will be useful to some extent for detecting chloroplast genome-length variability among accessions of *Nicotiana* or other Solanaceous plant species. It is possible to compare the average heterozygosity values for SSRs in coding and non-coding regions of the chloroplast genome. SSRs occurring within exons ($\hat{H}$=0.04, n=5) are considerably less variable than those occurring within introns ($\hat{H}$=0.29, n=8) or within intergenic regions ($\hat{H}$=0.38, n=19) of the chloroplast genome.

In this study, each accession from a wild or primitive cultivated species of potato has a unique haplotype. This result is, perhaps, not suprising since it is based upon the use of a single accession for each species. Given the robustness of these markers, it should be possible to develop multiplex assays whereby several of the cpSSRs are amplified concurrently in a single PCR reaction. Early experiments suggest that is it possible to develop PCR assays, whereby 4–5 of the loci described are amplified simultaneously (G. Bryan and J. McNicoll, unpublished results). This would further enhance the utility of these markers and provides a contrast with the difficulty in successfully multiplexing nuclear microsatellites.

Polymorphism at some chloroplast loci appears to be due to insertions and deletions of small segments of DNA, which are not directly associated with the mononucleotide repeats contained at the loci. A similar observation has been made for a human $(CA)_n$ microsatellite (Grimaldi and Crouau Roy 1997). These authors found that variations in the DNA flanking the SSR could generate size-homoplasy within species, despite the microsatellite repeats being of different length. Alleles of apparently identical size were found to have very different evolutionary histories. In the case of NTCP9 the insertion/deletion events into the flanking DNA are of a larger size than the variations in the SSR sequences, and, therefore, it is unlikely that events in the flanking DNA can produce apparent size homoplasies. The identification of highly polymorphic non-SSR tandem arrays suggests that it may be possible to find other polymorphic regions of the chloroplast genome that can be assayed using PCR. A second observation that cpSSRs deemed "too short" can show useful levels of polymorphism suggests a further source of polymorphic loci, should additional loci be required. Analysis of the *N. tabacum* chloroplast genome suggests that a further 26 loci may be available if the minimum criterion of Powell et al. (1995) is reduced to 9-bp rather than 10-bp. There appears to be no significant overall correlation between the level of polymorphism and the length of the repeat sequences for the 36 loci characterised thus far. However, a reduction in the minimum length of the SSR from 10 to 9-bp may bring about lower levels of polymorphism. Most of the variation in cpSSRs showing the "normal" pattern of variation is due to polymorphism in the length of the mononucleotide simple-sequence repeat. This suggests that it is, in general, appropriate to think of cpSSRs as microsatellites in the true sense and to analyse and interpret data obtained from them as such.

It is possible that cpSSRs can be used for aiding the species identification of germplasm material, for example that held within the Commonwealth Potato Collection. However, it is first necessary to perform further studies of within-species variability in cpDNA haplotypes. The results for the wild and primitive cultivated species themselves, although based upon the use of a single accession, suggest that these markers will be useful in the evaluation of material in collections of wild potato germplasm. Further studies based upon the use of more

accessions from each species should be carried out to determine to what extent species-specific or diagnostic cpSSR haplotypes can be established. A preliminary survey of accessions from the series Longipedicellata suggests that high levels of cpDNA diversity exist within wild species, and furthermore that there is considerable overlap between cpDNA haplotypes of accessions from closely related species, such as *S. papita* or *S. stoloniferum* (G. Bryan, unpublished data).

From an analysis of potato cpSSR haplotypes, it is clear that several potato cultivars possess haplotypes derived from other potato species, presumably as a result of attempts to widen the genetic base of cultivated potato or to introgress genes of agronomic importance, such as resistance to late blight or potato cyst nematode. The cultivars Brodick and Shelagh appear to have chloroplast genomes derived from Mexican species. Similarly, the breeding lines 62.33.3 and 65.346.19 possess haplotypes which also suggest that chloroplast genomes have been introgressed from wild species. Some of the wild species implicated are well-known for accessions showing resistance to late blight (e.g. *S. demissum*, *S. verrucosum*) and to potato cyst nematode (e.g. *S. vernei*). Other cultivars previously described to possess a chloroplast DNA haplotype derived from *S. tuberosum* ssp. *andigena* cluster more closely with a large proportion of the species examined in this study, including *S. tuberosum* ssp. *andigena*. This result may be partly explained by the very diverse nature of *S. tuberosum* ssp. *andigena*, and the use of a single accession of each species. These observations suggest that a considerable proportion of the chloroplast genome diversity present in cultivated potato is a direct result of the use of wild species as female parents. Our data also suggest that some of these cytoplasmic donors can be identified, given the use of the appropriate material for analyses.

A phylogeny has been constructed for the Solanaceae based upon chloroplast DNA restriction-site polymorphism (Olmstead and Palmer 1992). The inherently rapid rates of evolution of nuclear microsatellites have deemed them to be unsuitable for phylogenetic reconstruction. For example, Orti et al. (1997) compared nucleotide substitution data with SSR length variation and found a low correlation between size variation and genealogical relationships among alleles. However, given the purported slow rate of chloroplast DNA evolution, the markers described here would provide a useful test of the suitability of chloroplast genome-derived SSR markers for phylogenetic reconstruction in the Solanaceae.

Chloroplast DNA restriction-enzyme analysis has been used to address other questions concerning the origins of cultivated potato. Hosaka (1986) has investigated the possible maternal parent of the cultivated potato Chloroplast genome SSR markers, with their increased levels of resolution as compared to RFLP-based markers, should permit a reassessment of the question of the maternal origins of the common cultivated potato. These markers could also be used to test theories on the origins of allopolyploid species, such as *S. papita*, *S. stoloniferum* and *S. demissum*.

Acknowledgements We thank Jim McNicol for statistical advice and Jim Provan and Robbie Waugh for helpful discussions. This work was funded entirely by the Scottish Office Agriculture, Environment and Fisheries Department.

References

Avise, JC (1994) Molecular markers, natural history and evolution. Chapman and Hall, New York

Debener T, Salamini F, Gebhardt C (1990) Phylogeny of wild and cultivated *Solanum* species based on nuclear restriction fragment length polymorphisms (RFLPs). Theor Appl Genet 79:360–368

Grimaldi MC, Crouau Roy B (1997) Microsatellite allele homoplasy due to variable flanking sequences. Jour Mol Evol 44:336–340

Hawkes JG (1990) The potato – evolution, biodiversity and genetic resources. Belhaven Press, London

Hosaka K (1986) Who is the mother of potato? – restriction endonuclease analysis of chloroplast DNA of cultivated potatoes. Theor Appl Genet 72:606–618

Hosaka K, Hanneman RE (1988a) The origin of the cultivated tetraploid potato based on chloroplast DNA. Theor Appl Genet 76:172–176

Hosaka K, Hanneman RE (1988b) Origin of chloroplast DNA diversity in the Andean potatoes. Theor Appl Genet 76:333–340

Hosaka K, Ogihara Y, Matsubayashi M, Tsunewaki K (1984) Phylogenetic relationship between the tuberous *Solanum* species as revealed by restriction endonuclease analysis of chloroplast DNA. Jap J Genet 59:349–369

Murray MG, Thompson WF (1980) The isolation of high-molecular-weight plant DNA . Nucleic Acids Res 8:4321–4325

Nei M (1986) Molecular evolutionary Genetics. Columbia University Press, New York

Olmstead RG, Palmer JD (1992) A chloroplast DNA phylogeny of the Solanaceae: subfamilial relationships and character evolution. Ann Mis Bot Gard 79:346:360

Orti G, Pearse DE, Avise JC (1997) Phylogenetic assessment of length variation at a microsatellite locus. Proc Natl Acad Sci USA 94:10745–10749

Powell W, Baird E, Duncan N, Waugh R (1993) Chloroplast DNA variability in old and recently introduced potato cultivars. Ann Appl Biol 123:403–410

Powell W, Morgante M, McDevitt R, Vendramin GG, Rafalski JA (1995) Polymorphic simple-sequence-repeat regions in chloroplast: applications to the population-genetics of pines. Proc Natl Acad Sci USA 92: 7759–7763

Powell W, Morgante M, Andre C, McNicol JW, Machray GC, Doyle TS, Rafalski JA (1995) Hypervariable microsatellites provide a general source of DNA markers for the chloroplast genome. Curr Biol 5:1023–1029

Powell W, Morgante M, Doyle JJ, McNicol JW, Tingey SV, Rafalski AJ (1996) Genepool variation in genus *Glycine* subgenus *Soja* revealed by polymorphic nuclear and chloroplast microsatellites. Genetics 144:793–803

Provan J, Corbett G, Waugh R, McNicol JW, Morgante M, Powell (1996) DNA fingerprints of rice (*Oryza sativa*) obtained from chloroplast simple-sequence repeats. Proc R Soc Lond Series Sciences 263:1275–1281

Provan J, Powell W, Dewar H, Bryan G, Machray GC, Waugh R (1999) An extreme cytoplasmic bottleneck in the modern European cultivated potato (*Solanum tuberosum*) is not reflected in decreased levels of nuclear diversity. Proc Roy Soc Lond Series Sciences 266:633–639

Sneath, PHA, Sokal, RR (1973) Numerical taxonomy. Freeman, San Francisco, California

Soltis PS, Soltis DE, Doyle JJ (1992) Molecular systematics of plants. Chapman and Hall, New York

Spooner DM, Sytsma KJ, Conti E (1991) Chloroplast DNA evidence for genome differentiation in wild potatoes (*Solanum* sect. Petota: Solanaceae). Am J Bot 78:1354–1366