Theor Appl Genet (2000) 101:415–420

J. Lin · T.M. Davis

S1 analysis of long PCR heteroduplexes: detection of chloroplast indel polymorphisms in *Fragaria*

Received: 1 September 1999 / Accepted: 21 March 2000

Abstract S1 analysis of long PCR heteroduplexes was found to be an effective method for detecting phylogenetically informative indel (insertion/deletion) polymorphisms in the highly conserved strawberry chloroplast genome. In this broadly applicable method, long-range PCR products containing heteroduplex DNA molecules generated from mixed-template amplifications are subjected to S1 nuclease digestion followed by separation and visualization on an agarose gel. The presence of fragments resulting from S1 digestion of mismatch loops in heteroduplex molecules is indicative of indel polymorphism between the template sources. Upon analysis of 13-kb heteroduplex-containing PCR products spanning the *pet*A-*psb*B chloroplast genome region in *Fragaria vesca* and *Fragaria moschata*, two indels distinguishing these species were detected, and subsequently localized to the *psb*J-*psb*L and *rpl20*-*rps*18 intergenic regions. Comparative sequencing of these regions revealed that *F. moschata* resembled *Fragaria viridis*, but differed from *F. vesca*, *Fragaria nubicola*, and a closely related outgroup representative, *Duchesnea indica*, by a 10-bp deletion in the *psb*J-*psb*L region and a 10-bp insertion in the *rpl*20-*rps*18 region. Thus, of the three diploids (2n = 2*x* = 14) examined, *F. viridis* is favored over *F. vesca* and *F. nubicola* as the most likely chloroplast genome donor to hexaploid (2n = 6*x* = 42) *F. moschata*.

Key words *Fragaria* · Heteroduplex · Indel · Phylogenetics · Strawberry · S1 nuclease

Introduction

Indel (insertion/deletion) polymorphisms of potential value for phylogenetic and parentage studies have been identified in the chloroplast genomes of several plant species (Aldrich et al. 1988; Ogihara et al. 1988; van Ham et al. 1994). However, indels have not played a prominent role in phylogenetic analysis, in part because they are rare as compared with base substitutions (Saitou and Ueda 1994). Typically, indels have been detected as a by product of comparative sequencing. The value of indels as phylogenetic and analytical tools would be enhanced by the availability of a convenient means for their detection without sequencing. In ongoing studies of the *Fragaria* (strawberry) chloroplast genome, we have discovered several indel polymorphisms using a method we call S1 analysis of long PCR heteroduplexes. A similar approach was independently reported by Lundin et al. (1997) for detection of indels in the human mitochondrial genome.

The basic S1 nuclease technique for detecting indel polymorphisms was first introduced by Shenk et al. (1975). In this technique, homologous DNA fragments from different sources are mixed to form heteroduplex molecules, then treated with S1 nuclease. Because of its single-strand endonuclease activity, S1 nuclease cleaves heteroduplex molecules at sites of deletional mismatch. Thus, the presence of S1 fragments, as visualized on an electrophoretic gel, is diagnostic of one or more indel polymorphisms differentiating the two DNAs compared in the analysis. Non-PCR based S1 analysis has been performed using such substrates as cloned DNA fragments (Gannon et al. 1980), linearized plasmids (Burdon and Lees 1985), viral genomes (Silber and Loeb 1981), and whole chloroplast genomes (Serror et al. 1990). It has been shown that, although single-base-pair mismatch heteroduplexes are resistant to S1 nuclease digestion (Dodgson and Wells 1977; Silber and Loeb 1981), two-base and larger deletional mismatches are susceptible to S1 digestion (Shenk et al. 1975; Gannon et al. 1980; Burdon and Lees 1985). Thus, S1 heteroduplex analysis is a sensitive method of detecting indel polymorphisms.

Surprisingly, there are few reports of the application of S1 analysis to PCR product heteroduplexes. Tanaka-Yamamoto et al. (1989) employed S1 digestion of con-

Communicated by M.A. Saghai Maroof

J. Lin · T.M. Davis (✉)
Plant Biology Department, Rudman Hall,
University of New Hampshire, Durham, NH 03824 USA
e-mail: tom.davis@unh.edu

ventional PCR products to determine the location in the human mitochondrial genome of a deletion that was implicated in mitochondrial myopathy. In that application, S1 digestion was used, not to discover a deletional mutation, but to determine more precisely the location and extent of a suspected or known deletion. Lundin et al (1997) showed that S1 analysis of heteroduplexes generated from products of long-range PCR could be used diagnostically to detect a 9-bp mitochondrial deletion associated with human mitochondrial encephalomyopathies.

Briefly summarized, the S1 method works as follows. First, long-range PCR (Barnes 1994; Cheng et al. 1994) is used to amplify a sizable (approximately 5–20 kb) target region in template DNAs from two divergent sources. Long PCR-product heteroduplexes are generated, either by mixing separately amplified PCR products and then subjecting them to denaturation/renaturation (e.g., Lundin et al. 1997), or by co-amplification of the two templates in a mixed-template reaction as in the present study. The generation of heteroduplex molecules in mixed-template PCR occurs because in the late cycles of the PCR process an increasing proportion of the PCR-generated templates are not copied. Instead, many such templates (i.e., amplification products of previous cycles) reanneal with other such template molecules (Grompe 1993), thus forming a mixture of homoduplex and heteroduplex products. We have previously exploited this phenomenon in a screening procedure to identify primers that detect codominant RAPD (random amplified polymorphic DNA) markers (Davis et al. 1995).

Indels are detected by subjecting heteroduplex-containing PCR products to S1 nuclease digestion. S1 fragments are generated by nuclease digestion of the single strand loop that forms at an indel mismatch site and digestion of the alternate strand at the unpaired site opposite the digested loop (Shenk et al. 1975). The generation of S1 fragments in heteroduplex-containing products is diagnostic of the presence of one or more indels differentiating the template sources.

The genus *Fragaria* includes at least 15 species represented by four or more levels of ploidy (Hancock 1990; Staudt 1989). Efforts to establish phylogenetic relationships in *Fragaria* have been hampered by the low levels of variability detected in the chloroplast DNA of this genus (Harrison et al. 1997; Potter et al. 2000). In the present report, we describe the use of S1 analysis of long PCR heteroduplexes to detect phylogenetically informative indel polymorphisms in the chloroplast genomes of four *Fragaria* species: hexaploid (2n = 6*x* = 42) *F. moschata*, and diploids (2n = 2*x* = 14) *F. vesca*, *F. viridis* and *F. nubicola*.

Materials and methods

Four *Fragaria* species were represented in this study: the diploids *Fragaria viridis* (FRA333 and FRA 341), *Fragaria vesca* (FRA282 and 'Yellow Wonder'), and *Fragaria nubicola* (FRA520); and the hexaploid *Fragaria moschata* (FRA376, FRA157 and FRA609). Plants numbered with FRA prefixes were obtained as runner plants from the USDA Clonal Germplasm Repository, Corvallis, Oreg., while 'Yellow Wonder' was obtained as seed from W. Atlee Burpee and Co, Warminster, Pa. All of these strains are of European origin

Table 1 PCR primers and profiles. The primer prefixes L-, I-, and S- indicate that primers were used for long (L-), inverse (I-), or ''small'' (conventional) (S-) PCR, respecteively. Reaction tubes were heated to 94°C for 2 min before the first PCR cycle, and after the final cycle were held at 68°C for 10 min, then cooled to 5°C

Primer set	Primer sequence (5′–3′)	Primer location	Sequence source[a]	PCR profile[b]		Cycles
				Temp (°C)	Time	
Set 1						
L-PETA	ATCCCGACGGC (A/C) GCAAGAGAGTAA (C/T)A	*pet*A	T	D: 97.5	16 s	
				A: 59	1 min	35
L-PSBB	GCCAAAAAGCATAAGCCAGAAAAC	*psb*B	T	E: 68	10 min	
Set 2						
I-PETA1	TCGCCCTCGGAAACAAGAA	*pet*A	T	D: 97.5	16 s	
				A: 59	1 min	30
I-PETA2	TGGTCAGGGAGATGCAGAAATAGT	*pet*A	T	E: 68	2 min	
Set 3						
I-PSBB1	GGCCCAACCAGCAACCAGAG	*psb*B	T	D: 97.5	16 s	
				A: 63	1 min	30
I-PSBB2	ATGGGGCGGTTGGAGTATCACA	*psb*B	T	E: 68	5 min	
Set 4						
S-PSBJ	TGCCCGACTTTCCTCCACAT	5′-*psb*J	S	D: 94	30 s	
				A: 59	30 s	35
S-PSBF	GGCCTAGCTGTACCTACCGTCTTT	*psb*F	T	E: 68	30 s	
Set 5						
S-RPS18	TTAGTGAACAAGGAAAAATA	*rps*18	T	D: 94	30 s	
				A: 44	30 s	35
S-RPL20	TCATCGGGATAGGAGTAGA	*rpl*20	S	E: 68	40 s	

[a] T = tobacco, S = strawberry
[b] D= denaturation, A = annealing, E = extension

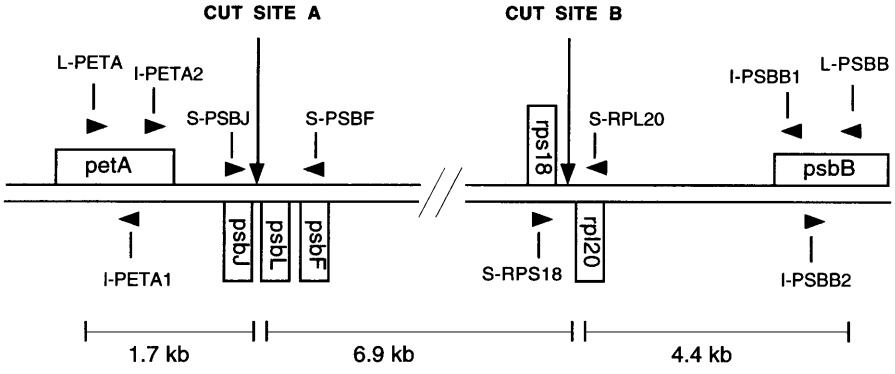

Fig. 1 Features of the *pet*A-*psb*B chloroplast DNA region. *Boxes* indicate gene locations and coding strands. Only relevant genes are shown. *Horizontal arrows* indicate primer sites. S1 cut-sites A and B are indicated by *vertical arrows*. S1 fragment positions and sizes are indicated at the bottom of the figure

except for FRA 520, from Pakistan. A representative of *Duchesnea indica*, of unknown geographic origin, was included because it is a close relative of *Fragaria* and is useful for outgroup comparison (Harrison et al. 1997).

Total strawberry DNA was isolated using the protocol of Davis et al (1995). A pair of PCR primers (Table 1, primer set 1) intended to amplify the strawberry chloroplast genome region bordered by the *pet*A and *psb*B genes (Fig. 1) were designed based upon known tobacco sequences (Shinozaki et al. 1986) using the program Lasergene Primer Select (DNAStar). Long-range PCR reactions were of 10 ul volumes containing 1 µl of 10 × high-salt buffer (with magnesium) and 0.1 µl of *Taq* Plus (Stratagene), 400 µM of each dNTP, 0.2 µM of each primer, 80 ng of template DNA, and dH_2O q.s. 10 µl. In mixed template reactions, 40 ng of template DNA from each of two sources was used. Before addition of template DNA, the otherwise complete reaction mix was vortexed for several seconds at medium speed. Each 0.5-ml tube was overlaid with 15 µl of sterile mineral oil. The long PCR profile is described in Table 1. Long PCR products were separated on a 0.8% Seakem (FMC BioProducts) agarose TBE gel run at 6 V/cm for 1.5 h. All gels were stained with ethidium bromide.

For use in S1 analysis, a long PCR product band containing homoduplex and (if present) heteroduplex DNA was excised and purified using the Geneclean Kit (Bio 101, Inc) according to the manufacturer's protocol, except that the gel was melted at 42°C rather than the prescribed 55°C to minimize apparent nicking of the double-stranded DNA. Using an S1 digestion protocol adapted from Sambrook et al. (1989), digestion conditions were empirically determined in an effort to maximize cutting at mismatch loops and minimize non-specific digestion. S1 digestions were done in 0.5-ml tubes in a reaction volume of 15 µl containing 12.5 µl of a long PCR Geneclean product (0.5–1.0 µg DNA), 1.5 µl of 10 × S1 buffer and 1 µl (5 units) of S1 nuclease (Promega). After digestion for 45 min at 20°C followed by 5 min at 45°C, the digestion was stopped by adding 1 µl of 0.4 M EDTA. S1 digestion products were separated on 1% agarose TBE gels (0.5% SeaKem and 0.5% NuSieve, FMC BioProducts) at 6 V/cm for 2 h. Because the pH of the S1 buffer was different from that of the DNA molecular-weight-marker, a blank lane was left between the marker and S1 digest lanes to avoid distortion of the marker bands.

To determine the sequence characteristics of S1 sites, selected S1 fragments were recovered from the gel using the previously described Geneclean procedure with 15 µl of elution solution, then circularized by self-ligation and subjected to inverse PCR using each of two primer pairs: primer set 2 situated within the *pet*A gene, and primer set 3 situated within the *psb*B gene (Table 1, Fig. 1). To protect the S1 fragments from nuclease degradation, TBE was used instead of water for gel staining and destaining. Each 15-µl ligation reaction containing 12.5 µl of S1 digestion fragment (50–100 ng of DNA), 1.5 µl of 10 × T4 ligase buffer and 1 ul (200 units) of T4 ligase (New England Biolabs) was incubated overnight at 16°C, then stored at –20°C for up to several months until used as a template for inverse PCR. The inverse PCR reaction mix was identical to the long PCR reaction mix previously described, but with 0.5 µl ligation solution used as a template. The inverse PCR products were then either directly sequenced at one end, using an inverse PCR primer as sequencing primer, or were cloned using the TA Cloning Kit (Invitrogen), then end-sequenced using vector primers.

Small PCR products containing regions immediately flanking two characterized S1 cut sites were amplified with a conventional *Taq* polymerase (Promega) reaction mix (as in Davis et al. 1995) using primer sets 4 and 5 (Table 1). To provide templates for sequencing, products of single template reactions were electrophoresed in 2% agarose TB E gels, excised and purified with the Geneclean procedure. For S1 analysis, single and mixed template PCR reaction products were digested by combining 25 µl PCR product with 2.9 µl of 10 × S1 buffer and 1 µl of S1 nuclease and incubating as described for S1 analysis above, followed by visualization on 2.5% agarose TBE gels.

Results

Single template long PCR amplifications using primer set 1 (Table 1) on *F. vesca* FRA282, *F. viridis* FRA333 and *F. moschata* FRA376 generated a single, approximately 13-kb product from each template (data not shown), as compared with a known size of about 10.3 kb for the *pet*A-*psb*B region in tobacco (Shinozaki et al. 1986). Direct end-sequencing of excised gel bands confirmed that the PCR product corresponded to the targeted *pet*A-*psb*B region. The effect of S1 treatment on the approximately 13-kb single template and mixed template PCR products from FRA282 and FRA376 was then determined, as visualized in Fig. 2. Three visible S1 fragments of sizes 1.7 kb, 4.4 kb and 6.9 kb were generated from the mixed template product of FRA282 + FRA376 (lane 7). Uncut homoduplex (and perhaps heteroduplex) products remain as a 13-kb band at the top of lane 7. No S1 fragments were evident below the 13-kb PCR products in the S1-treated FRA282 and FRA376 single template products (lanes 5 and 6, respectively), or in the untreated products (lanes 2 through 4). The sum of the three S1 fragment sizes, 1.7 + 4.4 + 6.9, equals 13-kb, the size of the long PCR product. The presence of three S1 fragments of a combined 13-kb size in the S1-treated, mixed template reaction products (lane 7) indicates that two sites within the 13-kb heteroduplex products were cut by S1 nuclease.

To locate and characterize the two S1 cut-sites, the three S1 fragments were recovered separately from the gel, circularized by self-ligation and subjected to PCR

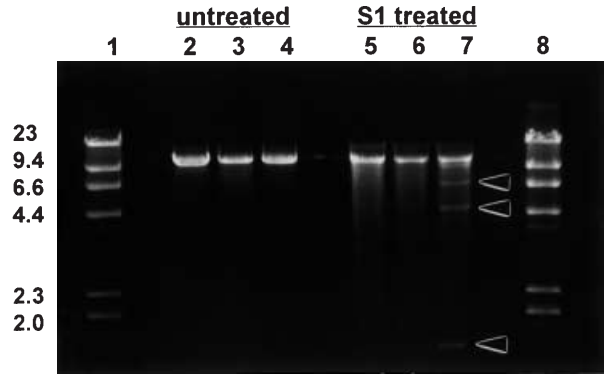

Fig. 2 S1 treatment of long PCR products. *Lanes 1 and 8*: Lambda *Hin*dIII molecular-weight marker (band sizes indicated at left in kb). PCR templates were FRA282 (*lanes 2 and 5*), FRA 376 (*lanes 3 and 6*) and a FRA282 + FRA376 mixed template (*lanes 4 and 7*). S1 treatment was applied to products in *lanes 5–7*, while *lanes 2–4* contain untreated controls. *Arrows* indicate characterized S1 fragments of 1.7-, 4.4- and 6.9-kb size in *lane 7*. An apparent S1 fragment of about 12 kb (unmarked) just below the uncut, homoduplex band in *lane 7* was not characterized

```
                          10         20         30         40
                          |          |          |          |
D. indica          AGAGAATTCC CACTATAAAA TACTAAATAT TAATAGTATA
'Yellow Wonder'    AGAGAATTCC TACTATAATA TACTATATAT TAGTAGTATA
FRA282             AGNGAATTCC TACTATAATA TAGTATATAT TAGTAGTATA
FRA520             AGAGAATTCC TACTATA-TA TACTATATAT TAGTAGTATA
FRA333             AGNGAATTCC TACTATA-TA ---------T TAGTAGTATA
FRA376             AGAGAATTCC TACTATA-TA ---------T TAGTAGTATA
```

Fig. 3 Alignment of sequences in the *psb*J-*psbF* intergenic region in *D. indica*, *F. vesca* ('Yellow Wonder' and FRA282), *F. nubicola* (FRA520), *F. viridis* (FRA333) *and F. moschata* (FRA376). The site A indel detected by S1 analysis of FRA282 and FRA376 is located at positions 18–27 (*dashes*)

```
                          10         20         30         40
                          |          |          |          |
D. indica          CGGATTCCTT TCAATCT--- -------CTT TATTTTATGA
'Yellow Wonder'    CAGATTCCTT TCACTCT--- -------CTT TATTTTATGA
FRA282             CAGATTCCTT TCACTCT--- -------CTT TATTTTATGA
FRA520             CAGATTCCTT TCACTCT--- -------CTT TATTTTATGA
FRA333             CAGATTCCTT TCACTCTCTT TATTTTACTT TATTTTATGA
FRA376             CAGATTCCTT TCACTCTCTT TATTTTACTT TATTTTATGA
```

Fig. 4 Sequence alignment in the *rps*18-*rpl*20 intergenic region in *D. indica*, *F. vesca* ('Yellow Wonder' and FRA282), *F. nubicola* (FRA520), *F. viridis* (FRA333) and *F. moschata* (FRA376). The site B indel detected by S1 analysis of FRA282 and FRA376 is located at positions 18–27 (*dashes*). The 10-base insertion at site B in FRA333 and FRA376 is an exact duplication of an adjacent 10-base sequence (*underlined* in FRA376) present in all six strains

amplification using inverse PCR primer sets 2 and 3 (Table 1). It was expected that each of these primer sets would yield a PCR product only when the circularized template contained the region targetted by the corresponding primer pair. Use of primer set 2 generated an approximately 1.7-kb product from the self-ligated 1.7-kb fragment, but generated no product from the self-ligated 4.4-kb and 6.9-kb fragments or from the three unligated fragments used as controls. Thus, the 1.7-kb S1 fragment must have been bounded by the *pet*A region at its uncut end, and the S1 cut-site (hereafter referred to as cut-site A) must have been 1.7-kb from the L-PETA long PCR primer site (Fig. 1). Similarly, use of primer set 3 generated an approximately 4.4-kb product from the self-ligated 4.4-kb fragment, but generated no product from the self-ligated 1.7-kb and 6.9-kb fragments or from the three unligated fragments used as controls. Thus, the 4.4-kb S1 fragment must have been bounded by the *psb*B region at its uncut end, and its S1 cut-site (hereafter referred to as cut-site B) must have been 4.4-kb from the L-PSBB long PCR primer site (Fig. 1). Because the 1.7- and 4.4-kb fragments were positioned at the *pet*A and *psb*B ends of the 13-kb PCR products, respectively, the 6.9-kb fragment, which did not amplify by inverse PCR using either of primer sets 2 or 3, must have been an internal fragment residing between the two S1 cut-sites.

Using primer I-PETA1 as a sequencing primer, about 400 bases of sequence were obtained from the 1.7-kb inverse PCR product. This sequence spanned the ligation site, and thus contained the sequence adjacent to (upstream of) cut-site A. However, when primer I-PSBB2 was used as a sequencing primer on the 4.4-kb inverse PCR product, the resulting sequence was good from the 5' end to the primer L-PSBB site but was garbled thereafter. To obtain the sequence adjacent to cut-site B, it was necessary to clone the 4.4-kb inverse PCR product and sequence the ends using vector sequencing primers. GenBank Blast searches revealed that sequenced regions adjacent to cut-sites A and B corresponded to the tobacco *psb*J and *rpl*20 genes, respectively. Using the known positions of the *pet*A, *psb*J, *rpl*20 and *psb*B genes within the tobacco chloroplast genome (Shinozaki et al. 1986) as a guide, and assuming conservation of chloroplast gene order between tobacco and strawberry, we deduced that S1 cut-site A was located in the region containing strawberry genes *psb*J-*psb*L-*psb*F, while cut-site B was between strawberry genes *rpl*20 and *rps*18 (Fig. 1).

Primer pairs (Table 1) that amplified the *psb*J-*psb*F (primer set 4) and *rpl*20-*rps*18 (primer set 5) intergenic regions were designed based upon our strawberry sequence data and known tobacco sequence. These primer pairs were used to amplify and obtain sequences from the *psb*J-*psb*F and *rpl*20-*rps*18 regions in FRA282 and FRA376. Sequence comparison in the *psb*J-*psb*F intergenic region revealed that FRA282 and FRA376 were distinguished by a 10-bp indel, evidently corresponding to S1 cut-site A, but were otherwise identical over the approximately 300-bp sequenced region (shown in part in Fig. 3). In the *rpl*20-*rps*18 intergenic region, FRA282 and FRA376 were also distinguished by a 10-bp indel, evidently corresponding to S1 cut-site B (Fig. 4). FRA282 and FRA376 were otherwise nearly identical over the approximately 360-bp sequenced region (shown in part in Fig. 4).

For comparison, the corresponding regions were also sequenced in *F. viridis* FRA333, *F. vesca*'Yellow Wonder', *F. nubicola* FRA520, and *D. indica*. In comparing the five *Fragaria* strains with the outgroup species *D. indica* at both indel sites, it was evident that FRA282, 'Yellow Wonder', and FRA520 resembled *D. indica* at

both sites, while FRA333 and FRA376 were distinguished from *D. indica* and the other *Fragaria* strains by a 10-bp deletion at site A (Fig. 3) and a 10-base insertion at site B (Fig. 4).

To confirm that the indels at sites A and B were potential sites of S1 digestion, the *psb*J-*psb*F and *rpl*20-*rps*18 regions were separately amplified with primer sets 4 and 5 in single-template and mixed-template reactions involving FRA282 and FRA376, and the products subjected to S1 digestion. For each primer set, heteroduplex products of FRA282 + FRA376 mixed-template reactions were cut by S1 into two fragments of the sizes predicted on the basis of the evident indel positions, while products of single-template (FRA282 or FRA376) reactions were uncut (data not shown).

Discussion

The goals of the research described here were to develop and evaluate a method of detecting indels, and to initiate a phylogenetic study of organelle genomes in the genus *Fragaria*. The method, S1 analysis of long PCR heteroduplexes, was successfully applied, resulting in the identification of two phylogenetically informative indel polymorphisms in a 13-kb region of the highly conserved strawberry chloroplast genome. We conclude that S1 analysis is a sensitive and effective means of detecting rare indel polymorphisms, with potential application to any genomic region of interest.

A key requirement for the application of S1 analysis described here is the initial availability of sequence information at potential PCR primer sites flanking an amplifiable region of interest. Organelle genomes are well-suited to this approach because the intergenic regions where indels are most likely to be present are relatively short. Thus, several such regions could be surveyed within a single, long PCR product generated using primers situated within flanking, conserved gene sequences. Another convenient target for S1 analysis would be introns of nuclear genes, in which indels could be sought by positioning primers in conserved exon regions flanking one or more intron sites.

S1 analysis can be applied to PCR products of any size. However, the likelihood of discovering rare indel increases in direct proportion to PCR product size, thus favoring the use of long-range PCR (Barnes 1994; Cheng et al. 1994). S1 analysis should detect indel polymorphisms occurring almost anywhere within the targeted sequence. Only polymorphisms within a few tens of base pairs of the ends of the long PCR products might be missed because the small, S1-generated end fragments might be too small to be detected on an agarose gel.

A possible concern about the use of mixed-template PCR to generate heteroduplex products for S1 analysis is that one of the two templates may fail to amplify, resulting in the absence of heteroduplexes and the subsequent failure to generate S1 fragments. Thus, if an indel did exist, the absence of S1 fragments would constitute a falsely negative result. This hazard can be avoided if single-template products are mixed post-PCR (Lundin et al. 1997), but only if product samples are checked on a test-gel to verify that individual PCR products are indeed present before mixing. However, in numerous positive control mixed-template reactions involving a variety of template sources and primers, we have never failed to detect heteroduplexes – either as diagnostic ‘‘non-parental’’ gel bands (Davis et al. 1995) or via S1 digestion – when the template sources were known to differ by a detectable indel. The mixed-template method has been dependable in our hands; therefore, we favor it because of its technical simplicity as compared with the post-PCR mixing method.

Depending on research goals, it may or may not be necessary to characterize S1 cut sites. If the goal is simply to detect a polymorphism useful as a genetic marker for inheritance, mapping or parentage studies, indel detection using S1 analysis of PCR products may be sufficiently informative. Alternately, if molecular characterization of the indel site is deemed necessary, the choice of strategy will depend on several factors, including the number and sizes of S1 fragments produced, the extent to which sequence information is available for primer design in the region of interest, and the suitability of circularized S1 fragments for inverse PCR and direct sequencing.

In the present study, comparative S1 analysis of *F. moschata* (FRA376) *and F. vesca* (FRA282) detected two indels within a 13-kb region of the chloroplast genome bounded by genes *pet*A and *psb*B. . Characterization of the S1 cut-sites showed that both indels were located in intergenic regions: site A between the *psb*J and *psb*L genes, and site B between the *rpl*20 and *rps*18 genes. Based upon sequence data obtained from regions flanking the cut-sites, primer pairs (sets 4 and 5) were designed to allow convenient amplification and sequencing or S1 analysis of small, conventional PCR products containing each indel individually.

Sequence comparison with outgroup *D. indica* provided a basis for inferring the ancestral and derived conditions at each indel site. At site A, *F. moschata* FRA376 had a 10-bp deletion (bases 82–91) as compared with *D. indica* and *F. vesca* FRA282 (Fig. 3). At site B, FRA376 had a 10-base insertion (bases 278–287) as compared with *D. indica* and FRA282 (Fig. 4). Thus, the allelic compositions at both sites are ancestral in *F. vesca* FRA 282 and derived in *F. moschata* FRA 376.

Indel sites A and B are located on the same DNA molecule; therefore, the multilocus genotype of an individual at these two sites constitutes a haplotype. If an insertion and a deletion are represented by the superscripts + and -, respectively, the haplotypes of the individuals in question can be represented as A-B+ for FRA376, and A+B- for *D. indica* and FRA282. The conservation of haplotype between *F. vesca* and the ‘‘outgroup’’ *D. indica* indicates that the A+B- haplotype is the ancestral state for *Fragaria*. The ancestral A+B- haplotype was also present in ‘Yellow Wonder’, a second representative

of European *F. vesca*, and in *F. nubicola* FRA520, an accession originally collected in Pakistan. *F. viridis* FRA333 and *F. moschata* FRA376, both of European origin, shared the derived A-B+ haplotype (as did *F. viridis* FRA 341 and *F. moschata* FRA157 and FRA609, data not shown)

These results provide insight into the phylogenetic relationship of hexaploid *F. moschata* to three potential diploid progenitors. The predominant diploid species within the geographic range of *F. moschata* are *F. vesca* and *F. viridis (*Staudt 1989). However, Staudt (1989) hypothesized that *F. moschata* might have arisen from the offspring of *F. vesca* × *F. nubicola* hybrids, based in part upon suggestive features of the latter hybrids. Among the three diploids examined here, our results favor *F. viridis* over both *F. vesca* and *F. nubicola* as the best candidate for the chloroplast donor to *F. moschata*. Final resolution of this issue must await the detection of additional genetic polymorphism and broader samplings of *Fragaria* germplasm.

Other than the two detected indels, little variation was observed within the sequenced regions among the four *Fragaria* species examined. Similarly, Potter et al. (2000) found little phylogenetically informative sequence variation in the *Fragaria* chloroplast *trn*L intron or the contiguous *trn*L-*trn*F spacer region. However, they did detect cpDNA mutations in *F. vesca and F. nubicola* that were lacking in *F. moschata,* consistent with our finding that neither of the former species were likely chloroplast DNA donors to *F. moschata*. In the only other report on *Fragaria* chloroplast genome polymorphism, an RFLP study of 26 *Fragaria* taxa by Harrison et al. (1997) detected insufficient variability to provide insight into the likely diploid progenitors of any of the polyploid species, including *F. moschata*. Overall, these studies show that the *Fragaria* chloroplast genome is highly conserved. However, our results promote the expectation that an expanded application of S1 analysis beyond of the *pet*A-*psb*B region will detect the additional indel polymorphisms needed for the resolution of a comprehensive *Fragaria* chloroplast phylogeny.

Acknowledgements This manuscript is scientific contribution number 2022 from the New Hampshire Agricultural Experiment Station. We thank Laura DiMeglio for obtaining the FRA520 DNA sequences, Jesse Davis for assistance with S1 analyses, and A.K. Klein, John Collins and Dan Potter for helpful reviews of the manuscript. The conduct of this research complied with current laws and regulations of the United States of America.

References

Aldrich J, Cherney BW, Merlin E, Christopherson L (1988) The role of insertions/deletions in the evolution of the intergenic region between *psb*A and *trn*H in the chloroplast genome. Curr Genet 14:137–146

Barnes WM (1994) PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. Proc Natl Acad Sci USA 91:2216–2220

Burdon MG, Lees JH (1985) Double-strand cleavage at a two-base deletion mismatch in a DNA heteroduplex by nuclease S1. Biosci Rep 5:627–632

Cheng S, Fockler C, Barnes WM, Higouchi R (1994) Effective amplification of long targets from cloned inserts and human genomic DNA. Proc Natl Acad Sci USA 91:5695–5699

Davis TM, Yu H, Haigis KM, McGowan PJ (1995) Template mixing: a method of enhancing detection and interpretation of codominant RAPD markers. Theor Appl Genet 91:582–588

Dodgson JB, Wells RD (1977) Action of single-strand specific nucleases on model DNA heteroduplexes of defined size and sequence. Biochemistry 16:2374–2379

Gannon F, Jeltsch JM, Perrin F (1980) A detailed comparison of the 5'-end of the ovalbumin gene cloned from chickpen oviduct and erythrocyte DNA. Nucleic Acids Res 8:4405–4421

Grompe M (1993) The detection of unknown mutations in nucleic acids. Nature Genet 5:111–117

Ham RCHJ van, Hart H, Mes THM, Sandbrink JM (1994) Molecular evolution of noncoding regions of the chloroplast genome in the Crassulaceae and related species. Curr Genet 25:558–566

Hancock, JF (1990) Ecological genetics of natural strawberry species. Hort Sci 25:869–871

Harrison RE, Luby JJ, and Furnier GR (1997) Chloroplast DNA restriction fragment variation among strawberry (*Fragaria* spp.) taxa. J Am Soc Hort Sci 122:63–68

Lundin K, Wilichowski E, Ernst B-P, Hanefeld F (1997) S1 nuclease hybrid analysis of mitochondrial DNA amplified by long-distance PCR: rapid screening for small-scale rearrangements. Nucleic Acids Res 25:2535–2536

Ogihara Y, Terachi T, Sasakuma T (1988) Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. Proc Natl Acad Sci USA 85:8573–8577

Potter D, Luby JJ, Harrison R (2000) Phylogenetic relationships among species of *Fragaria* (Rosaceae) inferred from noncoding nuclear and chloroplast DNA sequences. Systematic Bot 25:337–348

Saitou N, Ueda S (1994) Evolutionary rates of insertion and deletion in noncoding nucleotide sequences of primates. Mol Biol Evol 11:504–512

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Serror P, Hetraud F, Heizmann P (1990) Chloroplast DNA variability in the genus *Helianthus*: restriction analysis and S1 nuclease mapping of DNA-DNA heteroduplexes. Plant Mol Biol 15: 269–280

Shenk TE, Rhodes C, Rigby PWJ, Berg P (1975) Biochemical method for mapping mutational alterations in DNA with S1 nuclease: The location of deletions and temperature-sensitive mutations in Simian virus 40. Proc Natl Acad Sci USA 72:989–993

Shinozaki K, Ohme M, Tanaka M, et al. (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5:2043–2049

Silber JR, Loeb LA (1981) S1 nuclease does not cleave DNA at single-base mis-matches. Biochim Biophys Acta 656:256–264

Staudt, G (1989) The species of *Fragaria*, their taxonomy and geographical distribution. Acta Hort 265:23–33

Tanaka-Yamamoto T, Tanaka M, Ohno K, Satyo W, Horai S, Ozawa T (1989) Specific amplification of deleted mitochondrial DNA from a myopathic patient and analysis of deleted region with S1 nuclease. Biochimic Biophys Acta 1009:151–155