Plant Cell Rep (2015) 34:1681–1683
DOI 10.1007/s00299-015-1849-y

FOCUS

One size does not fit all: the risk of using amplicon size of chloroplast SSR marker for genetic relationship studies

Sun Woong Bang[1] · Sang-Min Chung[1]

Received: 11 June 2015 / Revised: 21 July 2015 / Accepted: 22 July 2015 / Published online: 8 August 2015

Key message **Even within closely related taxa, total length variation of PCR amplicons from chloroplast SSR must be confirmed by sequencing to avoid misinterpreting genetic relationships.**

Introduction and aims

In plant species, molecular analysis of the chloroplast genome has been used for investigating the evolution and systematics of genera and higher taxonomic ranks (Palmer et al. 1988). After chloroplast simple sequence repeat (cpSSR) markers were discovered, they were then developed and employed for genetic analyses. Doyle et al. (1998) addressed a potential risk of using length variation of SSR marker for genetic relationships, noting that the regions flanking SSRs could contain insertion or deletion (INDEL) or other SSR events. Despite warnings about size homoplasy, the length variation of cpSSR markers has been used in plants continuously for the study of genetic structure within closely related taxa (Karatas et al. 2014; Sanchez-Robles et al. 2014).

To reemphasize the risk in using length variation without sequence confirmation even within a plant species, we investigated the origin of total length variation of cpSSR within the species *Cucurbita pepo* L.

Communicated by N. Stewart.

✉ Sang-Min Chung
smchung@dongguk.edu

[1] Department of Life Science, Dongguk University-Seoul, Seoul 100-715, South Korea

Methodological approach

The seeds of five wild *C. pepo* accessions collected from Texas (TCN# 1135, 1139, and 1141) and from northeastern Mexico (TCN# 1156 and 1158) were kindly provided by Dr. Decker-Walter at The Cucurbit Network (TCN).

Twenty-three conserved chloroplast SSR (ccSSR) markers (ccSSR-1 to ccSSR-23) developed by Chung and Staub (2003) were employed in this study. Total DNA was extracted from one seed of each accession according to previously reported protocols (Chung et al. 2003). PCR amplification, DNA sequencing, and alignment were performed according to Chung and Staub (2003).

Findings

Among 23 ccSSR markers, all DNA sequences from ccSSRs using five *C. pepo* species had no polymorphisms, single-nucleotide substitutions, or single INDEL events for each amplified region of the ccSSRs (data not shown) excluding ccSSR-10. Marker ccSSR-10, located between chloroplast genes AtpB and RbcL, showed two INDEL events (two SSR motifs: T repeats and A repeats) in a single PCR amplified region (Fig. 1a).

Based on two SSR motifs, total length variations of TCN# 1135, 1139, 1141, 1156, and 1158 were 27 (16 T and 11 A), 27 (17 T and 10 A), 27 (17 T and 10 A), 28 (15 T and 13 A), and 28 (15 T and 13 A) base pairs, respectively. If the total size of PCR amplicons is only considered for genetic analysis with these samples, researchers would conclude that TCN# 1135, 1139, and 1141 shared genetic information in this cpDNA region. However, TCN# 1135 had 16 T repeats and 11 A repeats, while 1139 and 1141 had 17 T repeats and 10 A repeats, respectively (Fig. 1a).

A

```
1 C. pepo GAGTGAATTTTTTTTTTTTTTTT-ATTTCAATATTTTGAGGCAAAAGATTAGATTACAAANTTGAAAAAAAAAA--G
2 C. pepo GAGTGAATTTTTTTTTTTTTTTTTATTTCAATATTTTGAGGCAAAAGATTAGATTACAAANTTGAAAAAAAAA---G
3 C. pepo GAGTGAATTTTTTTTTTTTTTTTTATTNCAATATTTTGAGGCAAAAGATTAGATTACAAANTTGAAAAAAAAA---G
4 C. pepo GAGTGAATTTTTTTTTTTTTTTT--ATTTCAATATTTTGAGGCAAAAGATTAGATTNCAANNTTGAAAAAAAAAAAAG
5 C. pepo GAGTGAATTTTTTTTTTTTTTTT--ATTNCAATATTTTGAGGCAAAAGATTAGATTACAAANTTGAAAAAAAAAAAAG
```

B

```
1  E. elaterium    TAGAAAGCCTATTTTTTTT---AGTATTTATTAGCGAATTTGCTCTTTTTTTTT----CTTTCTATAGTG
2  D. palmatus     TAGAAAGCCTATTTTTTTTTTTAGTATTTATTAGCGAATTTGCTCTTTTTTTT-----CTTTCTATAGGG
3  C. palmate      TAGAAAGCCTATTTTTTTTT--AGTATTTATTAGCGAATTTGCTCTTTTTTTT-----CTTTCTATAGTG
4  B. dioica       TAGAAAGCCTATTGTTTTT---AGTATTTATTAGCGAAGTTGCTCTTTTTTTTTTTT-CTTTCTATAGGG
5  A. naudinianus  TAGAAAGCCTATTTTTTTT---AGTATTTATTAGCGAATTTGCTCTTTTTTTT-----CTTTCTATAGGG
6  C. colocynthis  TAGAAAGCCTATTTTTTTT---AGTATTTATTAGCGAATTTGCTCTTTTTTTT-----CTTTCTATAGTG
7  C. lanatus      TAGAAAGCCTATTTTTTTT---AGTATTTATTAGCGAATTTGCTCTTTTTTTT-----CTTTCTATAGTG
8  C. lanatus      TAGAAAGCCTATTTTTTTT---AGTATTTATTAGCGAATTTGCTCTTTTTTTT-----CTTTCTATAGGG
9  L. siceraria    TAGAAAGCCTATTTTTTTT--AGTATTTATTAGCGAATTTGCTCTTTTTTTT-----CTTTCTATAGGG
10 L. siceraria    TAGAAAGCCTATTTTTTTT--AGTATTTATTAGCGAATTTGCTCTTTTTTTT-----CTTTCTATAGTG
11 L. siceraria    TAGAAAGCCTATTTTTTTT--AGTATTTATTAGCGAATTTGCTCTTTTTTTT-----CTTTCTATAGTG
12 L. operculata   TAGAAAGCCTATTTTTTG----AGTATTTATTAGCGAATTTGCTCTTTTTTTTTT---CTTTCTATAGTG
13 L. cylindrica   TAGAAAGCCTATTTTTTG----AGTATTTATTAGCGAATTTGCTCTTTTTTTTTT---CTTTCTATAGGG
14 L. acutangula   TAGAAAGCCTATTTTTTT----AGTATTTATTAGCGAATTTGCTCTTTTTTTTTTT--CTTTCTATAGGG
15 L. quinquefida  TAGAAAGCCTATTTTTTG----AGTATTTATTAGCGAATTTGCTCTTTTTTTTTT---CTTTCTATAGGG
16 E. lobata       TAGAAAGCCTATTTTTTTT--AGNATTAATTAGGGAATTTGCTCTTTTTTTTT-----CTTTCTATAGAG
17 C. pedata       TAGAAAGCCTATTTTTTTT--AGTATTTATTAGTGAATTTGTTCTTTTTTGTTT---CTTTCTATAGNG
18 M. oreganos     TAGAAAGCCTATTTTTTTT---AGTATTGATTAGCGAATTTGCTCTTTTTTTTTTTTCTTCCAAAAGNG
19 B. hispida      TAGAAAGCCTATTTT-------AGTATTTATTAGCGAATTTGCTCTTTTTTTT-----CTTTCTATAGNG
20 B. hispida      TAGAAAGCCTATTTT-------AGTATTTATTAGCGAATTTGCTCTTTTTTTT-----CTTTCTATAGTG
21 C. pepo         TAGAAAGCCTATTTTTTTT---AGTATTTATTATCGAATTTGCTCTTTTTTTTTT---CTTTCTATAGNG
```

Fig. 1 **a** Alignment of ccSSR-10 fragment sequences from selected *C. pepo* L. in Cucurbitaceae. Insertion and/or deletion events are *bolded*. Uncertain nucleotides are denoted by *N*. *1* TCN (The Cucurbit Network) #1135, *2* TCN# 1139, *3* TCN# 1141, *4* TCN# 1156, *5* TCN# 1158. **b** Alignment of ccSSR-8 fragment sequences from selected genera of the Benincaseae tribe and outgroups in Cucurbitaceae (sequences were modified from Chung et al. 2003). Insertion and/or deletion events are *bolded*. Uncertain nucleotides are denoted by *N*

We confirmed from DNA sequence alignments that each INDEL event was generated by different genetic events, despite total length variations of the two INDEL events being identical. Consequently, the genetic information from each INDEL (i.e., 16 T repeats and 11 A repeats from TCN# 1135) must be treated as independent events (i.e., separation of two SSR motifs as two markers) for further genetic analyses. If the origin of these total length variations was not confirmed by sequencing, the genetic relationship based on size homoplasy would be misinterpreted and undoubtedly accepted by other researchers because the three *C. pepo* species, TCN# 1135, 1139, and 1141, are known to be very closely related taxa.

Although the development of ccSSR-10 was based on a single motif of 13 A repeats (Chung and Staub 2003), two SSR motifs (T repeats and A repeats) were detected in *C. pepo* in the targeted region. Therefore, we concluded that, depending on the taxa examined, there can be two SSR motifs (or INDEL) of chloroplast DNA even if cpSSRs are developed based on a single SSR motif.

DNA sequence alignment of ccSSR-10 indicated that two SSR motifs, 16 T repeats, and 11 A repeats are closely located, within 40 base pairs of each other (Fig. 1a). SSR primers are commonly selected and designed to target regions 50–100 base pairs upstream and downstream of the SSR motif. Therefore, if two SSR motifs are closely located (for example, less than 50 base pairs distance), the possibility to include two or more SSR motifs in a single SSR marker would be extremely high. In addition, two INDEL events (two SSR motifs) were found in a single cpSSR marker, ccSSR-8, located in the chloroplast gene Ycf3 (Fig. 1b) in our previous study (Chung et al. 2003). In this case, the distance between two INDEL events was only 21 base pairs. This leads to the expectation that SSR motifs in the chloroplast genome could be closely located depending on the taxa tested.

Although there have been numerous warnings about size homoplasy when using total length variation (Doyle et al. 1998; Wheeler et al. 2014), cpSSR markers have continuously been developed and employed for genetic analysis in plant species (Karatas et al. 2014; Sanchez-Robles et al. 2014; Phumichai et al. 2015). However, these studies did not report the origin of the total length variations in their data sets because the researchers probably believe that the use of closely related taxa (i.e., within a species) and a large number of markers reduced the risk for examining genetic relationship (Chung et al. 2003; Wheeler et al. 2014).

However, data from single misinterpreted polymorphic marker would generate a significant difference in a genetic relationship result when a small number of cpSSR markers are employed. Because our results showed a high rate of polymorphisms in cpSSR within a species, the polymorphic information from each SSR motif must be separated to obtain an accurate origin of the total length variation. Since two cpSSR motifs could be closely located (i.e., less than 20–50 base pairs), it would be challenging to separate each cpSSR motif for each SSR marker. Therefore, we conclude that even within a single species, results from genetic relationship studies must be reconsidered if total length variations in cpSSR were used without DNA sequence confirmation of the origin of total length variation.

Author contribution statement SB and SC designed research, conducted experiment, and wrote the manuscript. All authors read and approved the manuscript.

Acknowledgments This work was supported by grants from the Technology Development Program for Agriculture and Forestry, Ministry for Food, Agriculture, Forestry, and Fisheries (Grant Number 114078032HD020), Republic of Korea.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Chung S-M, Staub JE (2003) The development and evaluation of consensus chloroplast primer pairs that possess highly variable sequence regions in a diverse array of plant taxa. Theor Appl Genet 107:757–767

Chung S-M, Decker-Walters DS, Staub JE (2003) Genetic relationships within the Cucurbitaceae as assessed by ccSSR marker and sequence analyses. Can J Bot 81:814–832

Doyle JJ, Morgante M, Tingey SV, Powell W (1998) Size homoplasy in chloroplast microsatellites of wild perennial relatives of soybean (*Glycine* Subgenus *Glycine*). Mol Biol Evol 15:215–218

Karatas DD, Karatas H, Laucou V, Sarikamiş G, Riahi L, Bacilieri R, This P (2014) Genetic diversity of wild and cultivated grapevine accessions from southeast Turkey. Hereditas 151:73–80

Palmer JD, Jansen RK, Michaels HJ, Chase MK, Manhart JR (1988) Chloroplast DNA variation and plant phylogeny. Ann Mo Bot Gard 75:1180–1206

Phumichai C, Phumichai T, Wongkaew A (2015) Novel chloroplast microsatellite (cpSSR) markers for genetic diversity assessment of cultivated and wild *Hevea* rubber. Plant Mol Biol Rep. doi:10.1007/s11105-014-0850-x

Sanchez-Robles JM, Balao F, Terrab A, Garcia-Castano J, Ortiz MA, Vela E, Talavera S (2014) Phylogeography of SW Mediterranean firs: different European origins for the North African *Abies* species. Mol Phylogenet Evol 79:42–53

Wheeler GL, Dorman HE, Buchanan A, Challagundla L, Wallace LE (2014) A review of the prevalence, utility, and caveats of using chloroplast simple sequence repeats for studies of plant biology. Appl Plant Sci 2:1400059