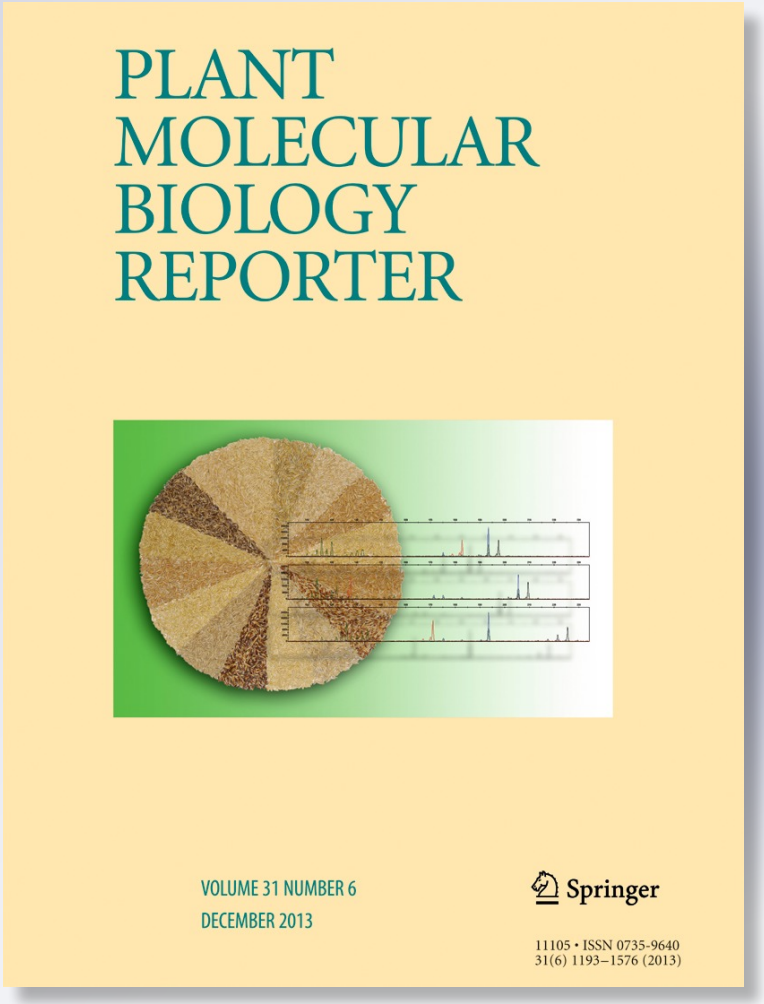
A Simple CELI Endonuclease-Based Protocol for Genotyping both SNPs and InDels

Guo-Qiang Jiang, Xue-Feng Yao & Chun-Ming Liu

Plant Molecular Biology Reporter

ISSN 0735-9640
Volume 31
Number 6

Plant Mol Biol Rep (2013) 31:1325-1335
DOI 10.1007/s11105-013-0606-z



Springer

Plant Mol Biol Rep (2013) 31:1325–1335
DOI 10.1007/s11105-013-0606-z

ORIGINAL PAPER

A Simple CELI Endonuclease-Based Protocol for Genotyping both SNPs and InDels

Guo-Qiang Jiang · Xue-Feng Yao · Chun-Ming Liu

Published online: 6 June 2013

Abstract Modern genetic analyses rely on efficient genotyping of single-nucleotide polymorphisms (SNP) or insertion/deletion length polymorphisms (InDel) in genomes. Methods available to genotype these polymorphisms include sequencing, cleaved amplified polymorphic sequence, high-resolution DNA melting, and microarray analyses, which are all rather tedious or expensive to set up for daily use. Here, we report a simplified label-free CELI endonuclease (CELI)-based protocol that enables us to detect both SNPs and InDels for fragment lengths between 500 and 6 kb. PCR-amplified target DNA fragments were annealed, cleaved by CELI, and analyzed either cost-effectively by agarose gel electrophoresis or automatically by capillary electrophoresis. The optimal amplification sizes, potential blind ends, and the maximum pooling capacities were examined for both electrophoresis protocols. We believe that the CELI-based genotyping protocol can be used in the detection of mutations, marker-assisted breeding, map-based cloning, and genome-wide association studies.

Keywords CELI · SNP · InDel · Agarose gel electrophoresis · Capillary electrophoresis

Abbreviations

CELI	CELI endonuclease
InDel	Insertion/deletion length polymorphism
MBN	Mung bean nuclease
SNP	Single-nucleotide polymorphism
TILLING	Targeting induced local lesions in genomes

Introduction

Single-nucleotide polymorphisms (SNP) and insertion/deletion length polymorphisms (InDel) are the two most frequently occurring genome variations, and they are frequently used in modern genetics for map-based cloning, linkage analysis, genome-wide association study, and marker-assisted selection. In humans, over 3.1 million SNPs have been found in haplotype maps, representing one SNP per 1 kb (International HapMap Consortium 2007; Wang et al. 2008). In rice, 1.7 million SNPs and 478,000 InDels have been found between sequenced genomes of Nipponbare (*Oryza sativa* L. ssp. *japonica*) and 93-11 (*O. sativa* L. ssp. *indica*), representing one SNP in every 268 bp and one InDel in every 953 bp of genomic DNA (Shen et al. 2004). Recently, more than 6.5 million SNPs were identified through re-sequencing 50 accessions of cultivated and wild rice (Xu et al. 2012).

A number of protocols have been developed for SNP and InDel detection with certain advantages and disadvantages associated with each of them. The cleaved amplified polymorphic sequence (CAPS; Konieczny and Ausubel 1993) and derived CAPS (dCAPS; Neff et al. 1998) methodologies are simple and effective but rely on the presence of restriction sites in either PCR-amplified fragments or in oligonucleotide primers. High-resolution DNA melting analysis is a rapid method established based on the loss of fluorescence when intercalating dyes are released from double-stranded DNA during thermal denaturation, but has the disadvantages that

Electronic supplementary material The online version of this article (doi:10.1007/s11105-013-0606-z) contains supplementary material, which is available to authorized users.

G.-Q. Jiang · X.-F. Yao · C.-M. Liu (✉)
Key Laboratory of Plant Molecular Physiology, Institute of Botany, Chinese Academy of Sciences, Nanxincun 20, Fragrant Hill, Beijing 100093, China
e-mail: cmliu@ibcas.ac.cn

G.-Q. Jiang
University of Chinese Academy of Sciences, 19A Yuquan Road, Beijing 100049, China

the target fragments analyzed are relatively short, and expensive equipment setup is required (Wittwer et al. 2003; Zhou et al. 2005; Montgomery et al. 2007; Li et al. 2010; Pereyra et al. 2012). Next-generation sequencing (NGS) and microarrays are ultra high-throughput technologies, allowing the simultaneous detection of thousands to millions of SNPs. However, genotyping a large number of individuals with NGS is still rather expensive (Gunderson et al. 2005; Nishida et al. 2008; Wang et al. 2008; Huang et al. 2010; Davey et al. 2011; Zheng et al. 2011; Chagné et al. 2012; Xu et al. 2012).

Targeting induced local lesions in genomes (TILLING) is a technology developed for detecting unknown point mutations in mutagenized populations based on CELI cleavage of mismatches (Colbert et al. 2001; Till et al. 2006; Wang et al. 2010). It has been used in many species including *Arabidopsis thaliana* (Colbert et al. 2001), *Lotus japonica* (Perry et al. 2003), maize (Till et al. 2004a), wheat (Uauy et al. 2009; Dong et al. 2009), tomato (Minoia et al. 2010), rice (Till et al. 2007), zebrafish (Wienholds et al. 2003; Sood et al. 2006) and fruit fly (Winkler et al. 2005). However, due to the complexity of the methodology, TILLING has not been used routinely in most laboratories for the detection of SNPs or InDels.

In this study, we modified the CELI-based technology for genotyping both SNPs and InDels using non-labeled primers in combination with either agarose gel or capillary electrophoresis. The minimum and maximum amplicon lengths, potential blind ends, and maximum pooling capacities were evaluated, which allowed us to develop a pair of label-free protocols for routine and high-throughput SNP and InDel detection.

Materials and Methods

Plant Materials and Genomic DNA Extractions

The protocol was optimized using three wild-type rice varieties: 93-11 (*O. sativa*, spp. *indica*), Nipponbare and Zhonghua 11 (*O. sativa*, spp. *japonica*), as well as 16 mutants in the Zhonghua 11 background. These mutants had point mutations in the following loci: *Os07g0301200*, *Os10g0389800*, and *Os06g0130100* (Table 1). Two DNA plasmids with one SNP (G to A) differing between them were used to evaluate the optimal length for amplicons. Primers used for SNP detection in this study are listed in Table 2. Five genomic loci with InDels of different sizes between Nipponbare and 93-11 were selected for evaluating the potential of using the CELI-based protocol for the detection of InDels. Genomic DNA was extracted from leaves of 1-month-old rice seedlings, as described by Wang et al. (2009).

PCR Reactions

Primer pairs were designed to produce amplicons of 0.7 to 9.7 kb in length. Ex Taq polymerase was used for amplification of fragments less than 2 kb in length (cat. no. DRR100B; TaKaRa, Japan). PCR with Ex Taq was performed with the following reaction conditions: 95 °C for 2 min; 8 cycles of 94 °C for 20 s, 68 °C for 30 s (1 °C decrease per cycle), and 72 °C for 60 s for every 1 kb of amplicon length, followed by 35 cycles of 94 °C for 20 s, 60 °C for 30 s, and 72 °C for 60 s for each 1 kb of amplicon length, and a final extension at 72 °C for 5 min. In amplifying fragments larger than 2 kb, KOD polymerase (cat. no. KFX-101; ToYoBo, Japan) was used in PCR reactions, and PCR was performed with the following reaction conditions: 95 °C for 2 min; 30 cycles of 98 °C for 20 s, 53 °C for 30 s, and 68 °C for 60 s for every 1 kb of amplicon length, followed by final extension at 68 °C for 10 min. To discriminate the genotype of individual plants, two types of reactions were tested and compared. The first one was to perform two parallel PCR reactions: one with DNA from a mutant and the second one with a 1:1 mixture of mutant and wild-type DNA. The second option was to perform one PCR reaction with DNA from the mutant and the other one with the wild-type DNA and then to mix equal amounts of PCR products from the mutant and wild-type reaction prior to annealing. All PCR products were subjected to a complete denaturation-slow annealing program to form heteroduplexes with the following conditions: 99 °C for 10 min for denaturation, followed by 70 cycles of decrements, starting at 70 °C, 20 s each, with a 0.3 °C decrease per cycle, and then holding at 15 °C to reanneal the denatured PCR products to form heteroduplexes.

CELI Extraction

CELI was extracted from 6 kg of celery (*Apium graveolens*) bought from a supermarket, as described by Till et al. (2006), with minor modifications. All steps were performed at 4 °C. After removing the white bottom and leaves, celery juice was collected using a prechilled juice extractor to process prechilled celery sticks. A buffer was added to the juice to make a final solution containing 100 mM Tris–HCl and 100 μM phenylmethylsulfonyl fluoride (PMSF), pH 7.7. After centrifugation at 2,600×*g* for 20 min, the supernatant was collected, and $(NH_4)_2SO_4$ was added gradually while stirring for at least 30 min to a final concentration of 25 % (m/*v*). After centrifugation at 15,000×*g* for 1 h, $(NH_4)_2SO_4$ was added gradually for at least 30 min to the supernatant while stirring to a final concentration of 80 % (m/*v*). The solution was then centrifuged at 15,000×*g* for 2 h, and the pellet was suspended with 100 mM Tris–HCl and 100 μM PMSF, pH 7.7, to one tenth of the starting volume of celery juice. Each 20 mL enzyme solution aliquot was dialyzed thoroughly in 4 L buffer containing 100 mM Tris–HCl, 500 mM KCl, and 100 μM PMSF, pH 7.7, for 1 h using dialysis bags with a cutoff size of 7 KDa. After four

Table 1 Rice mutants used to optimize CELI-based SNP detection

Gene loci	Line no.	Mutation[a]	Cleaved products (bp)[b]	Genotype
Os07g0301200 (1,489 bp amplicon)	1	G to A	141/1,348	Homo
	2	C to T	479/1,010	Hetero
	3	G to A	528/961	Hetero
	4	C to T	550/939	Homo
	5	G to A	565/924	Hetero
	6	G to A	613/876	Hetero
	7	G to A	635/854	Hetero
	8	G to A	638/851	Hetero
	9	No	No	Wild type
	10	G to A	749/740	Hetero
	11	C to T	798/691	Homo
	12	G to A	914/575	Homo
	13	C to T	1,221/268	Hetero
	14	C to T	1,229/260	Homo
	15	A to T	1,345/144	Homo
Os10g0389800 (1,220 bp amplicon)	16	G to A	890/330	Hetero
Os06g0130100 (1,427 bp amplicon)	17	G to A	938/489	Hetero

Hetero heterozygous, *Homo* homozygous

[a]Nucleotide in the wild type to the nucleotide in the mutant

[b]Sizes of the two expected CELI-cleaved products

replacements of the dialysis buffer, the enzyme solution was centrifuged at 10,000×*g* for 20 min, and the supernatant of CELI was stored at −80 °C.

CELI Digestion

CELI digestions of annealed PCR products were performed in 15 μL reaction mixture containing CELI buffer (10 mM HEPES, pH 7.5, 10 mM KCl, 10 mM $MgSO_4$, 0.002 % Triton X-100, and 0.2 μg/mL bovine serum albumin (BSA)), 4 μL of PCR product, and 1 unit CELI (10 units/μL) if PCR products were polymerized by Ex Taq (or 20 units CELI if PCR products were polymerized by KOD) at 45 °C for 15 min, followed by adding 3 μL of 0.5 M EDTA (pH 8.0) to stop the reaction. Alternatively, the digestion was performed in a 15-μL reaction mixture containing 4 μL of PCR products and 2 units of mung bean nuclease (MBN, 10 units/μL, cat. no. M0250S; New England Biolabs, USA) in MBN buffer (20 mM Bis-Tris, pH 6.5, 10 mM $MgSO_4$, 0.2 mM $ZnSO_4$, 0.002 % Triton X-100, and 0.2 μg/mL BSA) at 60 °C for 30 min, followed by adding 2 μL of 0.2 % SDS to stop the reaction.

Table 2 Primers used for CELI-based SNP detection

Primer name	Primer sequence (5′–3′)
Os07g0301200F	GCATGTGTATGTGCATGTCCACCTTTG
Os07g0301200R	TAAAGCTTTCGTCCTGTCATGCTGTGC
Os10g0389800F	AATTCCATGCAGTGACATGAGCCAAAA
Os10g0389800R	ACTGGTTGCGCTCAATCTTAGGAGGTG
Os06g0130100F	CACTCCATTTGCTTTTCTGCT
Os06g0130100R	GTCCACCGTCTTGGGCGTTG
Amplicon 9.7F	ATAAATGAGTAGATGAATTTTAG
Amplicon 7F	TACACTAGAAGGACAGTATTTG
Amplicon 6.8F	CTTAAAAATAATAAAAGCAGAC
Amplicon 6.1F	AACAATGGTGACTTCTAC
Amplicon 5.3F	GTTTTTAATGTACTGAATTAACG
Amplicon 4-2.7F	TCAGTCTCTTGTCGCTTAACG
Amplicon 1.9F	TAGATTTGCCATTGATTTGATACTC
Amplicon 1.9-2.7R	GTGTCTTTATTTTGCATTTTATACG
Amplicon 4-9.7R	TGATCAATTACTAACTACAATGG

Electrophoresis

CELI-digested PCR products were analyzed using 2 % agarose gels (cat. no. 9012-36-6; HydraGene, USA) containing 0.5 μg/mL ethidium bromide (cat. no. 0492; Amresco, USA) in 1× TAE buffer (40 mM Tris–acetic acid, 1 mM EDTA), with the voltage of 130 V for 60 min. Alternatively, CELI-digested PCR products in 96-well PCR plates (cat. no. AB19700; BIOplastics, The Netherlands) were diluted to 100 μL with deionized water, and capillary electrophoresis was performed at 9 kV, 30 s, for pre-run, 15 s for injection of 1 ng/μL molecular weight marker 75 and 15 kb or 50 and 3 kb dsDNA (Fermentas, Canada), 45 s for sample injection, and 40 min for sample separations in AdvanCE™ FS96 (Advanced Analytical Technologies, USA) by following the manufacturer's manual. Gel pictures were acquired and analyzed using the BioSpectrum AC Imaging System (UVP, USA) for agarose gels and PROSize software (Advanced Analytical Technologies, USA) for capillary electrophoresis.

Results

CELI Digestion and Agarose Gel (CELI/Agarose Gel-Based Genotyping) Protocol for SNP Detection

We applied CELI-based mismatch cleavage on reannealed, mixed PCR products containing a SNP and analyzed the cleavage products with agarose gel electrophoresis. This protocol is simple and efficient as compared to other methods. CELI extraction from celery is relatively simple and cost-effective, producing over 10,000 units in each preparation batch. Staining with ethidium bromide is more sensitive than other alternatives such as GoldView nucleic acid stain (SBS Bio, Beijing). To achieve optimal SNP detection, it is essential to set the PCR conditions properly to allow for specific amplification of target fragments. Attention must be given to primer design. When possible, all primers should be between 18 and 25 bp in length, their GC contents should be between 40 and 60 %, and their T_m values should be between 50 and 70 °C. It is often necessary to perform gradient PCR to identify an annealing temperature that gives a single PCR product with the expected size. Following PCR, denaturation-slow annealing, and CELI digestions, cleaved products were loaded on 2 % agarose gels and analyzed by electrophoresis. Increased concentrations of agarose to 2.5 or 3 % improved the detection of lower molecular weight bands but compromised the detection of high molecular weight bands (data not shown).

To examine if there were any biases in the detection of SNPs at different amplicon positions, we used a collection of 14 rice mutants, with point mutations throughout the 1,489-bp amplicon of the *Os07g0301200* locus (Table 1). All of these mutants were obtained by TILLING from an in-house EMS-mutagenized population made in the Zhonghua 11 (*O. sativa*, spp. *japonica*) background, with point mutations located on the 141st, 479th, 528th, 550th, 565th, 613rd, 635th, 638th, 749th, 798th, 914th, 1,221st, 1,229th, and 1,345th nucleotide in line nos. 1 to 8 and 10 to 15, respectively (Fig. 1a). Wild-type Zhonghua 11 (line 9) was used as a control. After evaluating point mutations located at different positions of PCR products, we noticed that it was difficult to detect cleaved fragments less than 250 bp, although the corresponding larger fragments in the same lane were visible (lines 1 and 15, Fig. 1). We therefore propose that for genotyping a known SNP locus, when possible, the primers should be designed to allow SNPs to be located 250 bp away from the two ends of the amplicon. We refer to these two terminal regions as "blind ends". As expected, the smallest amplicon for the CELI/agarose gel-based genotyping (CAG) protocol is around 500 bp. For genotyping unknown SNPs, we advise to have a 250-bp overlap between two adjacent PCR products, in order to obtain maximum coverage.

Optimization of the CAG Protocol

To detect homozygous mutations, individual DNA samples with point mutations were mixed either with an equal amount or with different ratios of wild-type DNA. After agarose gel electrophoresis, we observed two cleaved fragments for samples containing SNPs, in addition to the full-length uncleaved PCR products at the top of the gel (Fig. 1b, c). These two fragments, as expected, are a sum to the size of the full-length PCR products. For line no. 10, we noticed that only one cleaved fragment was observed, as the point mutation was located in the very middle of the PCR product (749th of 1,489 bp amplicon). Obviously, this is not ideal for effective SNP detection. We suggest that for analyzing SNPs at known positions, primers should be designed to prevent the SNP from being located in the very middle of the amplicon. Among different dilutions tested, we observed that all mutations were detected effectively up to a 1:7 mixing of the mutant and the wild-type DNA (Fig. 1c). As such, the low-cost agarose gel has about the same sensitivity for genotyping as the LI-COR 4300 sequencer reported previously (Colbert et al. 2001).

In the LI-COR 4300 sequencer-based method, the PCR products analyzed are usually no longer than 1.5 kb (Till et al. 2006). We evaluated the maximum amplicon size that could be analyzed by the CAG protocol. Two plasmids (with the total length of 14.7 bp) with one SNP (G to A) between them were mixed in a 1:1 ratio and used as the template. Primers were designed to produce SNP-containing amplicons with sizes of 1.9, 2.7, 4, 5.3, 6.1, 6.8, 7, or 9.7 kb (Fig. 2a). Following PCR, denaturation-slow annealing, and CELI digestions, samples were loaded on a 2 % agarose gel and analyzed by electrophoresis. As shown in Fig. 2b, cleaved fragments of the expected sizes were detected in all of these eight amplicons, in addition to the uncleaved PCR products located at the top of the gel. However, in lanes with total fragment lengths of 6.8, 7, and 9.7 kb, intensities of the expected 5.1, 5.4, and 8.0 kb cleaved products were difficult to distinguish from nonspecific bands. Since no other PCR fragments were observed in samples without CELI digestion, it is unlikely that they came from nonspecific amplification. Rather, these spurious bands likely resulted from nonspecific CELI digestion. As CEL1 is a single-strand endonuclease, non-perfect annealing of PCR products will lead to CELI cleavage at non-annealed single-strand regions. Such imperfect annealing does not seem to be a problem for PCR products less than 6 kb in length. As such, the upper detection limit for the CAG protocol should be considered to be about 6 kb. We sequenced more than 100 SNPs detected via CAG, and the accuracy rate was 100 %.

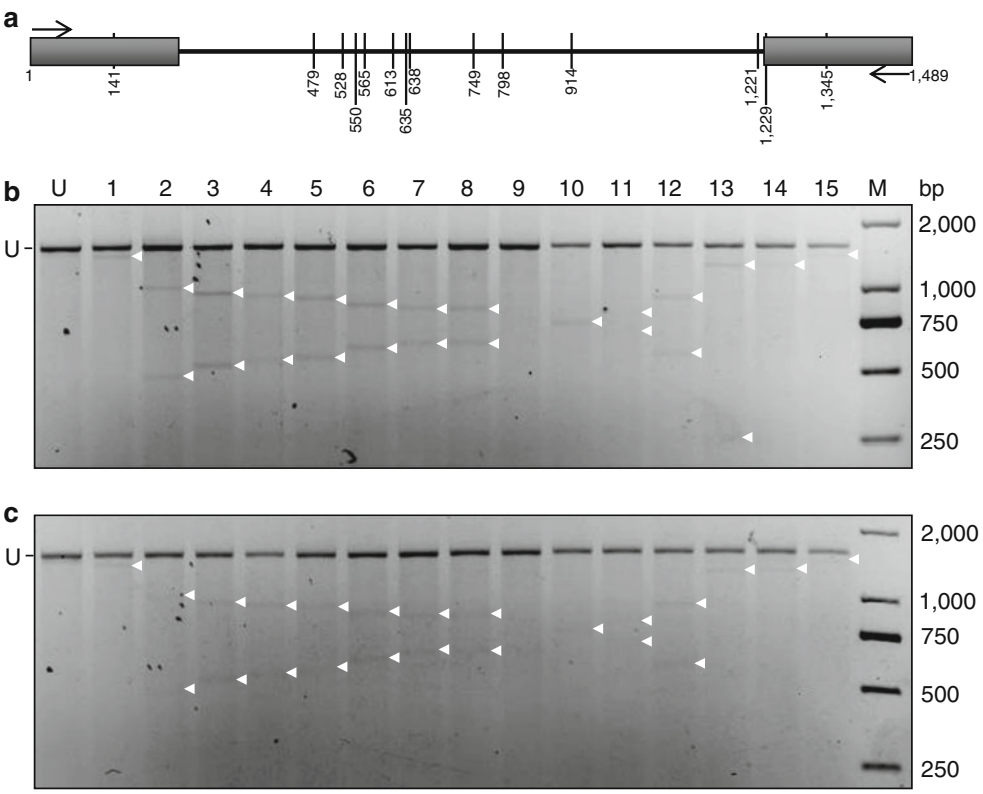


Fig. 1 Detection of SNPs at different positions using the CAG protocol. **a** Positions of different SNPs in the 1,489-bp amplicon of Os07g0301200. *Hatched boxes* correspond to the 250-bp blind ends. **b** A gel picture showing the cleaved PCR products (*arrowheads*) with 1:1 mixed mutant and wild-type DNA as the template. **c** A gel picture showing the cleaved PCR products (*arrowheads*) with a 1:7 mixture of mutant and wild-type DNA as the template. *U* uncleaved PCR products, *M* molecular weight marker

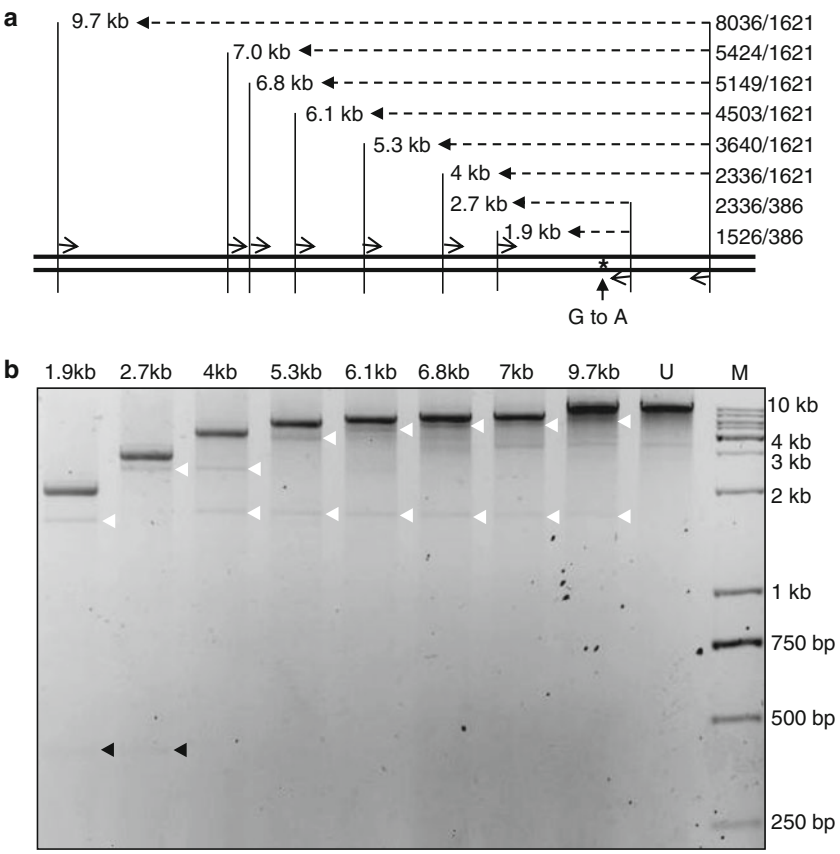


Fig. 2 Detection of SNPs in large amplicons using the CAG protocol. **a** Illustration of the positions of a particular SNP (*asterisk*) in different amplicons. *Numbers at the right* indicate the sizes of the expected cleaved PCR products. **b** A gel picture showing CELI-cleaved PCR products. The *dark band at the top of each lane* is the uncleaved PCR product. *Arrowheads* indicate CELI-cleaved products. Note that the 386 bp bands (*black arrowheads*) are barely visible. *U* PCR products without CELI digestion, *M* molecular weight marker

CAG-Based Genotyping in a Segregating Population

To ensure accurate detection of homozygous mutations, it is necessary to mix mutant and wild-type DNA. We used the CAG protocol to genotype 23 progeny produced from the self-pollinated heterozygous line no. 16 that carries a G-to-A point mutation in the *Os10g0389800* locus (Table 1). We designed a pair of primers to produce a 1.22-kb amplicon, with the point mutation located at the 890th nucleotide. Two independent sets of PCRs were performed, one with DNA from individual progeny plants and the second with equal amounts of wild-type DNA added. Following PCR, complete denaturation-slow annealing, and CELI digestion, these two sets of samples were loaded on 2 % agarose gels and analyzed independently. By comparing band patterns in these two gels, it was quite easy to define the genotype of all of the progeny plants (noted at the bottom of Fig. 3b). When no cleaved product was observed in either gel, the plant was a wild type (as in plants nos. 2, 3, 5, and 6). When cleaved products were observed in both gels, the plant was a heterozygous mutant (as in plant nos. 1, 7, 10, and 12). When cleaved products were observed only in the gels of the samples mixed with the wild-type DNA, the plant was a homozygous mutant (as in plant nos. 4, 8, 11, and 14). These three outcomes from the two gel strategies demonstrate that the CAG protocol is effective for genotyping SNPs in a segregating population.

CAG Protocol for InDel Analyses

To evaluate the possibility of using the CAG protocol for detecting InDels, we selected five genomic regions in which different sized InDels were present between the sequenced rice genomes of 93-11 and Nipponbare. As listed in Table 3, the sizes of InDel loci 1 to 5 were 33, 12, 11, 3, and 1 bp, respectively. We used the exact CAG protocol for SNPs as described above and observed that all of these InDels were detected effectively with agarose gel electrophoresis (Fig. 4). The sizes of the InDels did not affect cleavage efficiency. No difference was observed when CELI digestions were performed in PCR products from mixed DNA templates (Fig. 4, lane 2), nor in mixed PCR products from separated PCRs (Fig. 4, lane 3).

We also compared the difference between CELI and commercially available MBN for InDel cleavage reactions. Our results showed that while both MBN and CELI were able to cleave all five InDels (Fig. 4, lane 4 in each gel), the cleaved products were less abundant in the MBN samples than in the CELI samples. As with the CAG detection of SNPs, when InDels were located too close to the ends of a PCR product (e.g., 122 bp in InDel 5), the lower molecular weight band was more difficult to visualize (Fig. 5(e)).

CELI Digestion and Capillary Electrophoresis-Based Genotyping

Although the CAG protocol is cost-effective for detecting SNPs and InDels, dealing with a large number of samples using agarose gels is tedious and time-consuming. As an alternative, we developed a CELI/capillary electrophoresis-based genotyping protocol with a semiautomated approach using the AdvanCE™ FS96 instrument. The major advantage of the FS96 is able to perform analyses in 96-well samples automatically. Unlike the LI-COR system (Till et al. 2006), no post-digestion purification is needed with CELI/capillary electrophoresis-based genotyping (CCG). The mutants listed in Table 1 were used to optimize the CCG protocol (Fig. 5). To ensure parallel comparisons among samples in different capillaries, two molecular weight markers, 75 and 15 kb, were preloaded in each capillary. As shown in Fig. 5b, in addition to the full-length PCR products, cleaved PCR products were observed in all samples with point mutations, while no cleaved bands were seen in the wild-type controls. All expected cleaved

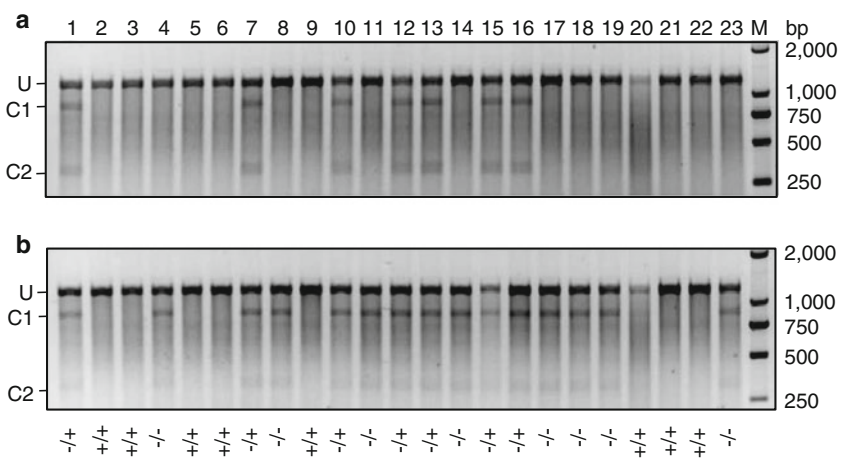


Fig. 3 Application of the CAG method to a segregating population. **a** A gel picture showing the CELI-cleaved PCR products of 23 progeny from a heterozygous line (no. *16*). **b** Same as in **a**, but the PCR templates were mixtures with equal amounts of both progeny and wild-type DNA. +/+, −/+, and −/− at the *bottom* indicate wild-type, heterozygote, or homozygote genotype, respectively. *U* uncleaved PCR products, *C1* and *C2* cleaved PCR products, *M* molecular weight marker

Table 3 InDel loci between 93-11 and Nipponbare used in this study

InDel locus no.	InDel sizes (bp)	Amplicon sizes (bp)	Cleaved products (bp)[a]	Primer sequence (5′–3′)
1	33	681	270/411	F: TCAATCGATAAATCCTTTGG
				R: CACTATGCCATACTAGAAG
2	12	831	387/432	F: TACGGATCCTGGGCGTG
				R: CATGGCATACATCAGAAAC
3	11	656	291/354	F: CCATGACCATTTTCACCC
				R: TTCCATCATTAATGCCTG
4	3	795	648/144	F: CCAAAATCTTCTTCATTGCA
				R: TATATATATTCTTCTTGTCC
5	1	804	122/681	F: TGAATGGGTAAACGGGAC
				R: CACATACCCGTTCAGGAG

F forward primer, *R* reverse primer

[a]Sizes of the two expected CELI-cleaved products

products, except the 141-bp product in line no. 1 (rather faint, Fig. 5), were effectively detected. We thus estimated that the "blind end" of CCG protocol was about 150 bp, which was approximately one half of the blind end size in the CAG protocol.

To assess the sensitivity of the CCG protocol, SNP-containing DNA samples were mixed in different ratios with wild-type DNA. We observed that SNPs were successfully detected in dilutions up to 1:23 (mutant/wild type) (Fig. 5c). This is dramatically more sensitive than the 1:7 dilution observed for the CAG protocol. We noticed that the 141 and 144 bp bands in line nos. 1 and 15 (indicated by black arrowheads, Fig. 5c) were rather faint, suggesting that the dilution limit had been reached.

The DNA samples and primers used to check the maximum detection length of the CCG protocol are shown in Fig. 2a. Although the SNPs in all eight amplicons were detected successfully, cleaved fragments in the amplicons with sizes of 6.8 kb or larger were difficult to distinguish from nonspecific bands (Fig. 6). As with CAG, the upper size limit for CCG can be considered to be about 6 kb.

The CCG protocol was then used to genotype a population of 22 progeny of the self-pollinated, heterozygous line no. 17 (Table 1), in which a G-to-A point mutation was located at the 938th nucleotide of the 1,427-bp amplicon from the *Os06g0130100* locus. Two independent sets of PCR were performed, one with DNAs from individual progenies and the second with equally mixed DNA from both individual progeny plants and the wild type. After CELI digestions, samples were analyzed by capillary electrophoresis using the FS96. By comparing band patterns between samples with and without mixing with wild-type DNA (Fig. 7b, a, respectively), the genotypes of each plant could be determined easily.

Detecting Multiple Mutations in a Sample

The CAG and CCG protocols described here differ from traditional LI-COR 4300 sequencer-based protocols in that they detect double, rather than terminal-labeled single-stranded DNA fragments. As such, they can, in principle, be used to identify more than one point mutation in a sample. We

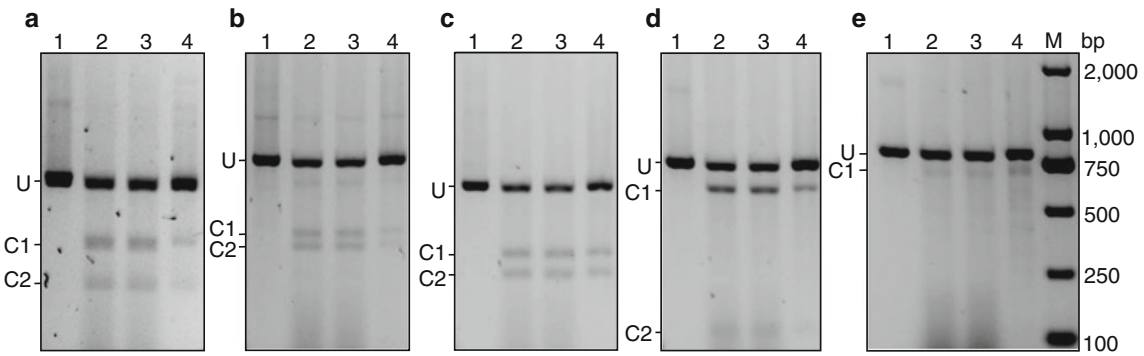


Fig. 4 Use of CAG for detecting InDels with different sizes using different endonucleases. *a–e* Gel pictures showing the detection of InDels with 33, 12, 11, 3, and 1 bp deletions, respectively. *Lane 1*, uncleaved PCR products; *lane 2*, CELI-cleaved PCR products, using equally mixed genomic DNA from Nipponbare and 93-11 as templates; *lane 3*, CELI-cleaved equally mixed PCR products, using separate genomic DNA of Nipponbare or 93-11 as the template; *lane 4*, MBN-cleaved equally mixed PCR products. *U* uncleaved PCR products, *C1* and *C2* cleaved PCR products, *M* molecular weight marker

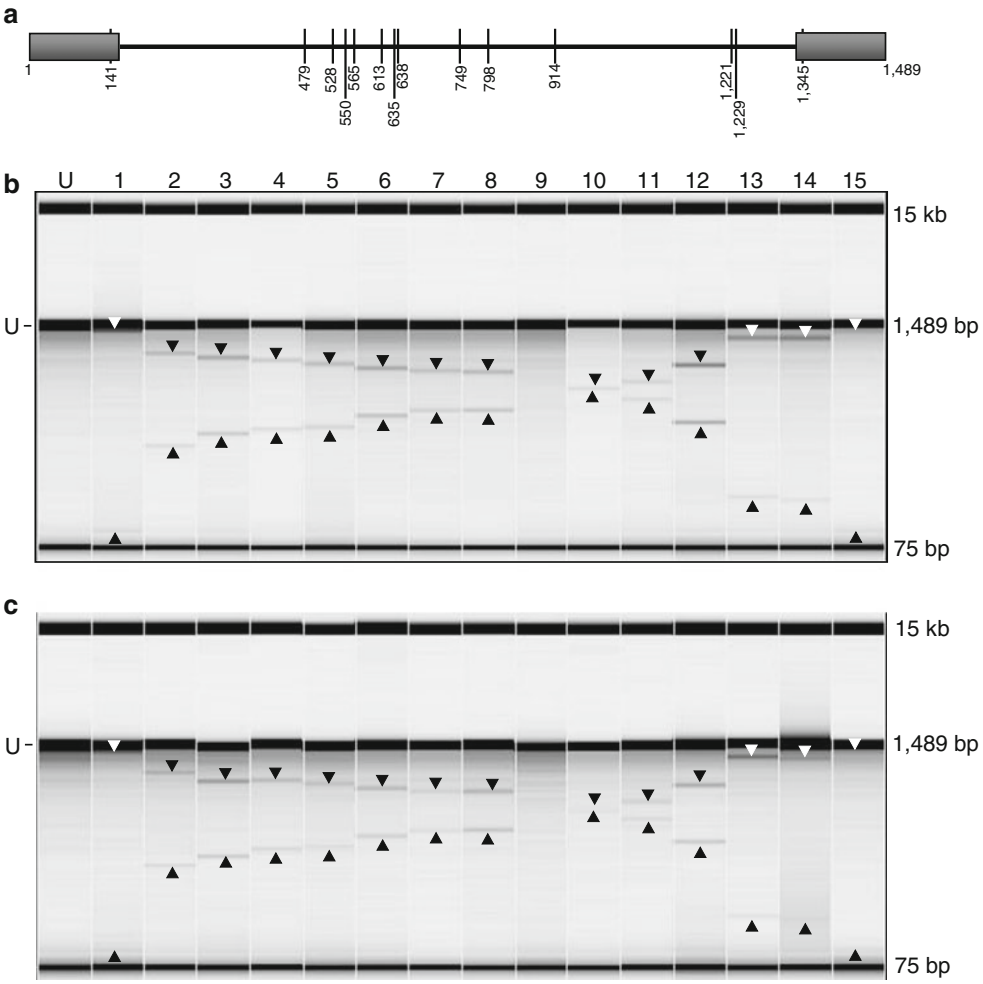


Fig. 5 Detection of SNPs at different positions using the CCG protocol. **a** Positions of different SNPs in the 1,489-bp amplicon of Os07g0301200. *Hatched boxes* correspond to the 150-bp blind ends. **b**, **c** Pseudogel pictures of capillary electrophoresis showing the cleaved PCR products (*arrowheads*) with mixed mutant and wild-type DNA at 1:1 and 1:23 ratios, respectively. The *top* and *bottom bands* are 15 kb and 75 bp markers, respectively. *U* uncleaved PCR products

examined if this was indeed the case using PCR products with two different point mutations. As shown in Fig. 8, both the CAG and CCG protocols were able to detect samples with two mutations, showing four bands (Fig. 8a) or peaks (Fig. 8b) in addition to the full-length uncleaved band.

Discussion

CELI is a single-strand specific endonuclease. The enzyme cuts one of the two DNA strands in a mismatched duplex (Oleykowski et al. 1998; Desai and Shankar 2003). By

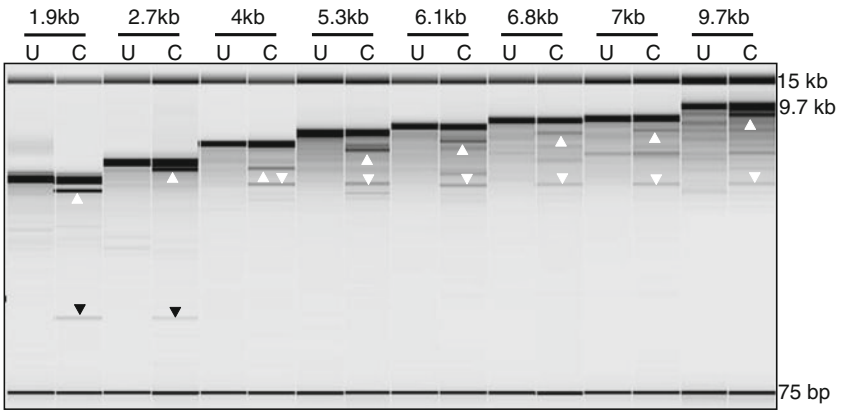


Fig. 6 Detection of SNPs in large amplicons using the CCG protocol. Positions of SNP in different amplicons and the sizes of the expected cleaved PCR products are shown in Fig. 2a. The *top* and *bottom bands* are 15 kb and 75 bp markers, respectively. The *second band from the top of each lane* is the uncleaved PCR product. *Arrowheads* indicate CELI-cleaved products. Note that the 386-bp fragments (*black arrowheads*) are evidently visible. *U* uncleaved PCR products, *C* cleaved PCR products

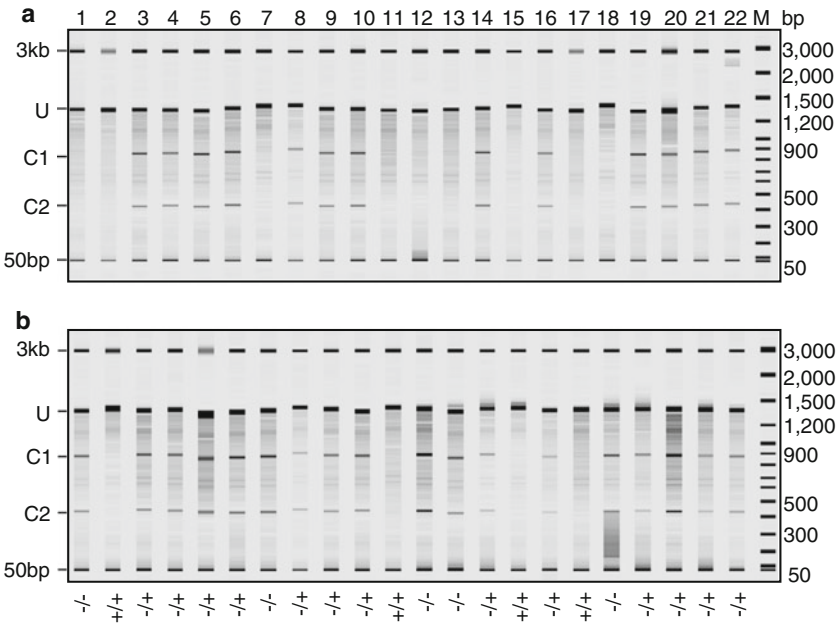


Fig. 7 Application of the CCG protocol to a segregating population. **a** A pseudogel picture of capillary electrophoresis showing the CELI-cleaved PCR products of 22 progeny from a heterozygous line (no. *17*). **b** Same as in **a**, but the PCR templates were mixtures with equal amounts of both progeny and wild-type DNA. +/+, −/+, and −/− at the *bottom* indicate wild-type, heterozygote, or homozygote genotype, respectively. *U* uncleaved PCR products, *C1* and *C2* cleaved PCR products, *M* molecular weight marker

increasing either the enzyme concentration or incubation time, CELI can also cut both strands at the mismatch site (Sokurenko et al. 2001). CELI has been applied in TILLING for detection of mutations using florescence-labeled primers

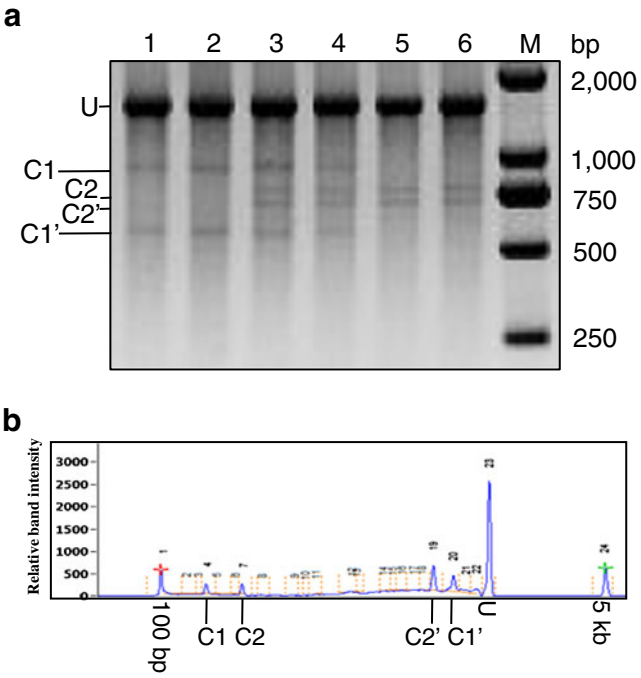


Fig. 8 Detecting two point mutations in one sample using the CAG (**a**) and CCG (**b**) protocols. **a** Samples *1* and *2* contain SNP nos. 1 (924/580 bp), and *3* and *4* contain both SNP nos. 1 and 2, while *5* and *6* contain SNP no. 2 (784/720 bp). *U* uncleaved PCR product of 1,504 bp; *M* molecular weight marker. **b** Peak pattern in a sample with two different SNPs, showing two pairs of peaks (C1 and C1′, C2 and C2′) of cleaved PCR products in addition to the uncleaved full-length PCR (*U*). The "100 bp" and "5 kb" are molecular weight markers

in combination with the LI-COR 4300 sequencer (Colbert et al. 2001; Perry et al. 2003; Till et al. 2004b, 2006). Due to the complexity of the procedure, CELI-based SNP detection has not been used routinely in most laboratories. In this study, we developed and optimized two CELI-based protocols (CAG or CCG) which are able to analyze both SNP and InDel mutations without using labeling primers. The optimal amplification lengths and detection regions for amplicons were evaluated for both protocols.

The various SNP and InDel detection protocols all have both advantages and problems. Agarose gel-based CAPS or dCAPS protocols have been used extensively to genotype SNPs but require the presence of restriction sites in the amplified fragments or in the primers (Konieczny and Ausubel 1993; Neff et al. 1998). Simple sequence length polymorphism markers are easy to use in analyzing InDels (Panaud et al. 1996), although InDels less than 10 bp in length are often difficult to detect. Measuring the melting points of heteroduplexes is another method to detect both SNP and InDel mutations, but the drawback is that only fragments of relatively short length (100–300 bp) can be analyzed (Wittwer et al. 2003; Reed and Wittwer 2004). In recent years, high-throughput sequencing has been used widely in the detection of both SNP and InDel mutations at the whole-genome level. However, the method can be too complicated for small labs or research programs in areas lacking these technologies. Additionally, use of sequencing for analysis of a large population of individuals can be prohibitively expensive (Wang et al. 2008; Deschamps et al. 2010; Huang et al. 2010; Davey et al. 2011; Zheng et al. 2011; Xu et al. 2012; Subbaiyan et al. 2012). Errors resulting through the use of different calling algorithms are another problematic shortfall of sequencing-based SNP and

InDel detection methods. For example, among 224 randomly selected SNPs identified between two chickpea genotypes from Illumina sequence data, only 50 % were confirmed to be real mutations following validation (Azam et al. 2012). The traditional TILLING protocol with the LI-COR 4300 sequencer can be used for SNP detection but requires florescent dye-labeled primers, which is both time-consuming and expensive (Colbert et al. 2001; Till et al. 2006). The CAG and CCG protocols introduced in this paper use ordinary primers and are able to detect SNPs and InDels. The use of non-labeled primers is critically important, as it saves tremendous amounts of time and money. As for efficiency, the CAG protocol is able to detect SNPs and InDels in 8X pooled samples in a cost-effective way. The CCG protocol is able to detect mutations in 24X pooled DNA samples in 96-well plates automatically, with a greatly increased sensitivity and throughput, as compared to the traditional LI-COR 4300 system.

The CAG and CCG protocols have additional advantages. Firstly, as SNPs and InDels are the two most frequently occurring genomic variations in nature (Shen et al. 2004), effective genotyping of both of these mutation types with the same methodology provides an efficient tool for genetic analyses and map-based cloning. Secondly, both of these protocols are able to analyze large fragments with total length up to 6 kb. Thirdly, the CCG protocol does not require the purification of CELI-digested PCR products, allowing for the analysis of 96 samples in less than 1 h.

In summary, this paper introduces a pair of CELI-based label-free protocols for genotyping both SNP and InDel mutations (see the complete protocols in 1 and 2 of the Electronic Supplementary Material). In small-scale experiments, we advise the use of the simple CAG protocol because of its convenience and cost-effectiveness. For genotyping a large number of samples (a hundred to thousands), the semiautomated CCG protocol is a better choice due to its higher throughput and sensitivity. We believe that these two protocols will have many applications in molecular genetic studies.

Acknowledgments We acknowledge the supports from the National Transgenic Research Projects (2011ZX08009-001-006-002; 2011ZX08009-003-005) and EU CEDROME project (INCO-CT-2005-015468).

References

Azam S, Thakur V, Ruperao P, Shah T, Balaji J, Amindala B, Farmer AD, Studholme DJ, May GD, Edwards D, Jones JD, Varshney RK (2012) Coverage-based consensus calling (CbCC) of short sequence reads and comparison of CbCC results to identify SNPs in chickpea (*Cicer arietinum*; Fabaceae), a crop species without a reference genome. Am J Bot 99:186–192

Chagné D, Crowhurst RN, Troggio M, Davey MW, Gilmore B, Lawley C, Vanderzande S, Hellens RP, Kumar S, Cestaro A, Velasco R, Main D, Rees JD, Iezzoni A, Mockler T, Wilhelm L, Van de Weg E, Gardiner SE, Bassil N, Peace C (2012) Genome-wide SNP detection, validation, and development of an 8 k SNP array for apple. PLoS One 7:e31745

Colbert T, Till BJ, Tompa R, Reynolds S, Steine MN, Yeung AT, McCallum CM, Comai L, Henikoff S (2001) High-throughput screening for induced point mutations. Plant Physiol 126:480–484

Consortium IHM (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449:851–861

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12:499–510

Desai NA, Shankar V (2003) Single-strand-specific nucleases. FEMS Microbiol Rev 26:457–491

Deschamps S, Rota M, Ratashak JP, Biddle P, Thureen D, Farmer A, Luck S, Beatty M, Nagasawa N, Michael L, Llaca V, Sakai H, May G, Lightner J, Campbell MA (2010) Rapid genome-wide single nucleotide polymorphism discovery in soybean and rice via deep resequencing of reduced representation libraries with the Illumina genome analyzer. Plant Genome 3:53–68

Dong C, Dalton-Morgan J, Vincent K, Sharp P (2009) A modified TILLING method for wheat breeding. Plant Genome 2:39–47

Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS (2005) A genome-wide scalable SNP genotyping assay using microarray technology. Nat Genet 37:549–554

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang QF, Li J, Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961–967

Konieczny A, Ausubel FM (1993) A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. Plant J 4:403–410

Li YD, Chu ZZ, Liu XG, Jing HC, Liu YG, Hao DY (2010) A cost-effective high-resolution melting approach using the EvaGreen dye for DNA polymorphism detection and genotyping in plants. J Integr Plant Biol 52:1036–1042

Minoia S, Petrozza A, D'Onofrio O, Piron F (2010) A new mutant genetic resource for tomato crop improvement by TILLING technology. BMC Research Notes 3:69

Montgomery J, Wittwer CT, Palais R, Zhou L (2007) Simultaneous mutation scanning and genotyping by high-resolution DNA melting analysis. Nat Protoc 2:59–66

Neff MM, Neff JD, Chory J, Pepper AE (1998) dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in *Arabidopsis thaliana* genetics. Plant J 14:387–392

Nishida N, Koike A, Tajima A, Ogasawara Y, Ishibashi Y, Uehara Y, Inoue I, Tokunaga K (2008) Evaluating the performance of Affymetrix SNP Array 6.0 platform with 400 Japanese individuals. BMC Genomics 9:431

Oleykowski CA, Bronson Mullins CR, Godwin AK, Yeung AT (1998) Mutation detection using a novel plant endonuclease. Nucleic Acids Res 26:4597–4602

Panaud O, Chen X, McCouch SR (1996) Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). Mol Gen Genet 252:597–607

Pereyra S, Velazquez T, Bertoni B, Sapiro R (2012) Rapid multiplex high resolution melting method to analyze inflammatory related SNPs in preterm birth. BMC Res Notes 5:69

Perry JA, Wang TL, Welham TJ, Gardner S, Pike JM, Yoshida S, Parniske M (2003) A TILLING reverse genetics tool and a web-accessible collection of mutants of the legume *Lotus japonicus*. Plant Physiol 131:866–871

Reed GH, Wittwer CT (2004) Sensitivity and specificity of single-nucleotide polymorphism scanning by high-resolution melting analysis. Clin Chem 50:1748–1754

Shen YJ, Jiang H, Jin JP, Zhang ZB, Xi B, He YY, Wang G, Wang C, Qian L, Li X, Yu QB, Liu HJ, Chen DH, Gao JH, Huang H, Shi TL, Yang ZN (2004) Development of genome-wide DNA polymorphism database for map-based cloning of rice genes. Plant Physiol 135:1198–1205

Sokurenko EV, Tchesnokova V, Yeung AT, Oleykowski CA, Trintchina E, Hughes KT, Rashid RA, Brint JM, Moseley SL, Lory S (2001) Detection of simple mutations and polymorphisms in large genomic regions. Nucleic Acids Res 29:E111

Sood R, English MA, Jones M, Mullikin J, Wang DM, Anderson M, Wu D, Chandrasekharappa SC, Yu J, Zhang J, Paul Liu P (2006) Methods for reverse genetic screening in zebrafish by resequencing and TILLING. Methods 39:220–227

Subbaiyan GK, Waters DL, Katiyar SK, Sadananda AR, Vaddadi S, Henry RJ (2012) Genome-wide DNA polymorphisms in elite indica rice inbreds discovered by whole-genome sequencing. Plant Biotechnol J 10:623–634

Till BJ, Reynolds SH, Weil C, Springer N, Burtner C, Young K, Bowers E, Codomo CA, Enns LC, Odden AR, Greene EA, Comai L, Henikoff S (2004a) Discovery of induced point mutations in maize genes by TILLING. BMC Plant Biol 4:12

Till BJ, Burtner C, Comai L, Henikoff S (2004b) Mismatch cleavage by single-strand specific nucleases. Nucleic Acids Res 32:2632–2641

Till BJ, Zerr T, Comai L, Henikoff S (2006) A protocol for TILLING and Ecotilling in plants and animals. Nat Protoc 1:2465–2477

Till BJ, Cooper J, Tai TH, Colowit P, Greene EA, Henikoff S, Comai L (2007) Discovery of chemically induced mutations in rice by TILLING. BMC Plant Biol 7:19

Uauy C, Paraiso F, Colasuonno P, Tran RK (2009) A modified TILLING approach to detect induced mutations in tetraploid and hexaploid wheat. BMC Plant Biol 9:115

Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, Fan W, Zhang J, Li J, Zhang J, Guo Y, Feng B, Li H, Lu Y, Fang X, Liang H, Du Z, Li D, Zhao Y, Hu Y, Yang Z, Zheng H, Hellmann I, Inouye M, Pool J, Yi X, Zhao J, Duan J, Zhou Y, Qin J, Ma L, Li G, Yang Z, Zhang G, Yang B, Yu C, Liang F, Li W, Li S, Li D, Ni P, Ruan J, Li Q, Zhu H, Liu D, Lu Z, Li N, Guo G, Zhang J, Ye J, Fang L, Hao Q, Chen Q, Liang Y, Su Y, San A, Ping C, Yang S, Chen F, Li L, Zhou K, Zheng H, Ren Y, Yang L, Gao Y, Yang G, Li Z, Feng X, Kristiansen K, Wong GK, Nielsen R, Durbin R, Bolund L, Zhang X, Li S, Yang H, Wang J (2008) The diploid genome sequence of an Asian individual. Nature 456:60–65

Wang L, Long Y, Liu Y (2009) A method for rapid preparation of plant genomic DNA for PCR analysis. Mol Plant Breeding 7:425–428

Wang T, Uauy C, Till B, Liu CM (2010) TILLING and associated technologies. J Integr Plant Biol 52:1027–1030

Wienholds E, van Eeden F, Kosters M, Mudde J, Plasterk RH, Cuppen E (2003) Efficient target-selected mutagenesis in zebrafish. Genome Res 13:2700–2707

Winkler S, Schwabedissen A, Backasch D, Bokel C, Seidel C, Bonisch S, Furthauer M, Kuhrs A, Cobreros L, Brand M, Gonzalez-Gaitan M (2005) Target-selected mutant screen by TILLING in *Drosophila*. Genome Res 15:718–723

Wittwer CT, Reed GH, Gundry CN, Vandersteen JG, Pryor RJ (2003) High-resolution genotyping by amplicon melting analysis using LCGreen. Clin Chem 49:853–860

Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, Dong Y, Gutenkunst RN, Fang L, Huang L, Li J, He W, Zhang G, Zheng X, Zhang F, Li Y, Yu C, Kristiansen K, Zhang X, Wang J, Wright M, McCouch S, Nielsen R, Wang J, Wang W (2012) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat Biotechnol 30:105–111

Zheng LY, Guo XS, He B, Sun LJ, Peng Y, Dong SS, Liu TF, Jiang S, Ramachandran S, Liu CM, Jing HC (2011) Genome-wide patterns of genetic variation in sweet and grain sorghum (*Sorghum bicolor*). Genome Biol 21:R114

Zhou L, Wang L, Palais R, Pryor R, Wittwer CT (2005) High-resolution DNA melting analysis for simultaneous mutation scanning and genotyping in solution. Clin Chem 51:1770–1777