Mol Breeding
DOI 10.1007/s11032-010-9462-9

Single nucleotide polymorphism genotyping by heteroduplex analysis in sunflower (*Helianthus annuus* L.)

Corina M. Fusari · Verónica V. Lia · Verónica Nishinakamasu · Jeremías E. Zubrzycki · Andrea F. Puebla · Alberto E. Maligne · H. Esteban Hopp · Ruth A. Heinz · Norma B. Paniego

Received: 15 February 2010 / Accepted: 13 May 2010

Abstract Single nucleotide polymorphisms (SNPs) and insertions/deletions (indels) are increasingly used for cultivar identification, construction of genetic maps, genetic diversity assessment, association mapping and marker-assisted breeding. Although there are several highly sensitive methods for the detection of polymorphisms, most of them are often beyond the budget of medium-throughput academic laboratories or seed companies. Heteroduplex analysis by enzymatic cleavage (CEL1CH) or denaturing high-performance liquid chromatography (dHPLC) has been successfully used to examine genetic variation in several plant and animal species. In this work, we assess and compare the performance of both methods in sunflower by genotyping SNPs from a set of 24 selected polymorphic candidate genes. The CEL1CH method allowed us to accurately detect allele differences in 10 out of 24 regions using an in-house prepared CEL1 enzyme (celery single strand endonuclease 1, *Apium graveolens* L.). Similarly, a total of 11 regions were successfully optimized for dHPLC analysis. As a scaling-up approach, both strategies were tested to genotype either 42 SNPs/indels in 22 sunflower accessions from the local germplasm bank or 33 SNPs/indels in 90 recombinant inbred lines (RILs) for genetic mapping purposes. Summarizing, a total of 601 genotypes were efficiently analyzed either with CEL1CH (110) or dHPCL (491). In conclusion, CEL1CH and dHPLC proved to be robust, complementary methods, allowing medium-scale laboratories to scale up the number of both SNPs and individuals to be included in genetic studies and targeted germplasm diversity characterization (EcoTILLING).

Keywords Sunflower · SNPs · Heteroduplex analysis · CEL1 · dHPLC · High-throughput genotyping

Electronic supplementary material The online version of this article (doi:10.1007/s11032-010-9462-9) contains supplementary material, which is available to authorized users.

C. M. Fusari · V. V. Lia · V. Nishinakamasu · J. E. Zubrzycki · A. F. Puebla · A. E. Maligne · H. E. Hopp · R. A. Heinz · N. B. Paniego (✉)
Instituto Nacional de Tecnología Agropecuaria (INTA), Instituto de Biotecnología (CNIA), CC 25, Hurlingham, Buenos Aires 1686, Argentina
e-mail: npaniego@cnia.inta.gov.ar

V. V. Lia · H. E. Hopp · R. A. Heinz
Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Intendente Güiraldes 2160, Ciudad Universitaria 1428, Argentina

Introduction

Single nucleotide polymorphisms (SNPs) and insertions/deletions (indels) are the most abundant types of DNA polymorphisms. Their use is linked with many applications such as cultivar identification,

Published online: 06 June 2010

construction of genetic maps, assessment of genetic diversity and linkage disequilibrium, association mapping and marker-assisted breeding (Lijavetzky et al. 2007). Furthermore, the development of high-throughput genotyping methods makes SNPs highly attractive as genetic markers. Although there are several methods for SNP genotyping (Comai et al. 2004; Lin et al. 2009; Makridakis and Reichardt 2001; Ren et al. 2004; Tsuchihashi and Dracopoli 2002; Wang et al. 2005; Xiao and Oefner 2001), most of them are often beyond the budget of low-to-medium-throughput academic laboratories and breeding programs. In order to limit the number of markers that need to be typed, many plant biotech laboratories are pursuing population and association mapping studies by candidate gene rather than genome-wide approaches (Dracatos et al. 2008; Ehrenreich et al. 2009; Pajerowska-Mukhtar et al. 2009). Those projects demand simple, accurate and cost-effective technology for SNP genotyping on either individuals or pooled DNA samples.

SNPs analyses based on the recognition of heteroduplex DNA molecules provide an efficient tool for examining genetic variation, and different methodological approaches are presently available, including oligonucleotide hybridization (Pease et al. 1994; Southern 1996), Taq-Man® assays (Higuchi et al. 1993; Livak et al. 1995), molecular beacons (Barreiro et al. 2009; Tyagi and Kramer 1996; Wang et al. 2009), electronic dot blot assays (Gilles et al. 1999), denaturing high-performance liquid chromatography (dHPLC) (Oefner and Underhill 1998) and specific endonuclease cleavage approaches (Oleykowski et al. 1998). Genotyping methods based on heteroduplex analysis typically compare two or more alleles (alleles A and B, Fig. 1) as a mixture of denatured and re-annealed polymerase chain reaction (PCR) amplicons. When differences in nucleotide sequence are found between the alleles, two types of double-stranded DNA molecules are again formed upon re-annealing: (1) the original homoduplex molecules (Homo, Fig. 1) and (2) heteroduplex molecules composed of the sense and antisense strands of either allele (Hetero, Fig. 1).

CEL1CH (heteroduplex analysis by enzymatic cleavage) has been successfully used in TILLING (Targeting Induced Local Lesions IN Genomes) to identify point mutations derived from ethylmethanesulfonate-mutagenized populations from different plant species (Cooper et al. 2008; Perry et al. 2003; Till et al. 2007; Uauy et al. 2009; Wang et al. 2008; Xin et al. 2008) and to assess natural variation (EcoTILLING) (Comai et al. 2004; Galeano et al. 2009; Gilchrist et al. 2006; Nieto et al. 2007). CEL1 (celery single strand endonuclease 1, *Apium graveolens* L.) is a family member of plant endonucleases that recognizes mismatches in heteroduplex DNA and cleaves both strands on the 3′ side of the mismatch distortion (Fig. 1). Mutation detection with CEL1 enzyme is available through commercial kits optimized for different platforms, including WAVE® Systems, standard gel electrophoresis, fluorescent capillary electrophoresis and LI-COR (Surveyor Mutation Kit, Transgenomic Inc., USA). Nevertheless, the enzyme can be obtained from celery stalks, making this method less expensive for low-budget laboratories.

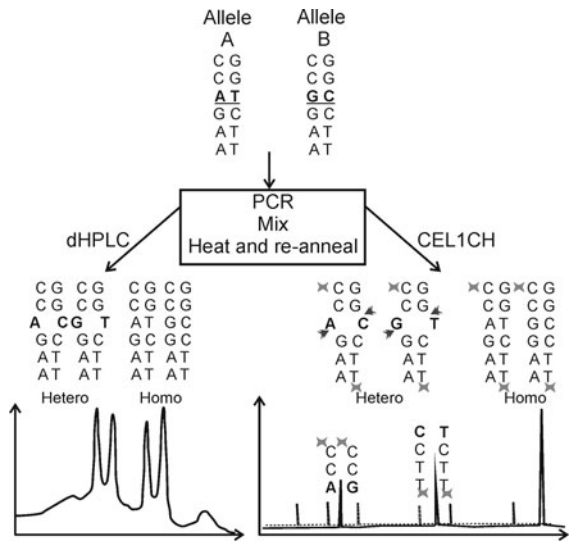

Fig. 1 Outline of SNP detection by heteroduplex analysis. Amplified regions of different alleles (allele A and B) are mixed in equimolecular proportions and subjected to a heating and cooling process to enable the formation of homoduplex and heteroduplex molecules. In dHPLC, heteroduplex molecules elute earlier than the homoduplex because of their reduced melting temperature. In CEL1CH, a DNA automatic analyzer enables the detection of labeled fragments corresponding to cleaved heteroduplex and homoduplex molecules

Alternatively, dHPLC reveals the presence of genetic variation by the differential retention of the homo- and heteroduplex DNA molecules on ion-pair reversed-phase high-pressure liquid chromatography supported under partial denaturing conditions (Xiao and Oefner 2001). In a typical assay, heteroduplex

molecules elute from the column prior to the homoduplex molecules, because of their reduced melting temperature (Fig. 1). The dHPLC method has been demonstrated to be the superior technique for SNP detection in terms of sensitivity and efficiency, although the cost of purchasing specific dHPLC equipment may be a limiting factor for academic laboratories (Randall et al. 2005). This method has been successfully and extensively used in human diagnosis to detect mutations involved in a number of diverse diseases (Giordano et al. 1999; Wagner et al. 1999) and to assess antibiotic resistance mutations in bacteria and microbial communities (Randall et al. 2005). dHPLC application in plants has been reported in the assessment of natural variation (Spiegelman et al. 2000), candidate gene analysis (Lexer et al. 2003) and SNP mapping (Kota et al. 2008, 2001; Lai et al. 2005).

In sunflower, direct sequencing of several genomic regions in a small group of inbreds and landraces has yielded a diverse array of functional markers (ca. 1,729 SNPs and 147 indels) that could potentially be exploited for germplasm characterization and association mapping studies (Fusari et al. 2008; Kolkman et al. 2007; Liu and Burke 2006). The aim of this work was to assess the performance of CEL1CH and dHPLC techniques in genotyping SNPs in sunflower. Optimization of these methods will provide the sunflower community with suitable tools to expand genotyping efforts to a wider range of inbred lines, cultivars and landraces and to add functional markers to sunflower linkage maps.

Materials and methods

Plant material

Nine elite sunflower inbred lines, previously sequenced for 28 candidate genes (Fusari et al. 2008), were employed as controls for the evaluation and optimization of two methodologies based on heteroduplex formation. These inbred lines—HA89 (PI 599773, CGGI00787), HA61 (PI 599771), HA292 (PI 552937, CGGI01061), HAR3 (PI 650754), HAR5 (PI 650763), KLM280 (CGGI0098), PAC2 (kindly provided by Dr. Gentzbittel, EnSat, France), RHA266 (PI 599755, CGGI00377M) and RHA801 (PI 599768, CGGI00982)—were chosen from the original group of 19 inbreds based on their genotypes to allow representation of the haplotypes found for the tested candidate genes. A second set of sunflower inbred lines was used to assay CEL1CH and dHPLC performance for the analysis of individuals of unknown genotype. These 22 germplasm accessions were MP557 (CGGI00436), HA336 (CGGI00374), F164B (CGGI00608), PAC1 (CGGI00494), P94 (CGGI00607), CM307 (CGGI00610), L156 (CGGI00631), HA64 (PI 599772, CGGI01063), HAR4 (PI 650755, CGGI00803), RK416 (CGGI1099), RHA275 (PI599760), HA853 (NSL 202855, CGGI00849B), HA441 (PI 639164), RHA440 (PI 639163, CGGI00827), RHA439 (PI 639162), RHA274*b* (PI 599759, CGGI00841), R432, A71, A59, R419, R417 and R423 (kindly provided by Dr. Daniel Alvarez who carried out the INTA pre-breeding program). All germplasm accessions belong to the Germplasm Bank of the Estación Experimental Agropecuaria INTA Manfredi (Córdoba, Argentina) (http://www.inta.gov.ar/balcarce/banco_germop/solicitud.htm). They carry interesting agronomic characteristics including resistance to fungal diseases, tolerance to abiotic stress, increased number of seeds per capitulum and high oil yield. In addition, a marker subset was mapped using 90 recombinant inbred lines (RILs) derived from a cross between PAC2 and RHA266. This RIL population had previously been used in a number of studies for genetic and quantitative trait locus (QTL) mapping (e.g. Al-Chaarani et al. 2002, 2004, Alfadil et al. 2007, Poormohammad Kiani et al. 2007a, b, 2008). Both parental genotypes are sunflower public inbred lines. RHA266 was obtained from a cross between wild *H. annuus* and Peredovik by USDA and PAC2 is an INRA-France inbred line from a cross between *H. petiolaris* and HA61. RHA266 is more resistant to downy mildew with higher values for yield, 1000-grain weight and oil percentage compared with PAC2. PAC2 is more resistant to *Sclerotinia sclerotiorum* (Poormohammad Kiani et al. 2007b).

Candidate genes and primer design

A group of 24 candidate genes previously characterized by Fusari et al. (2008) was selected to evaluate CEL1CH and dHPLC methods. The IDs, putative functions inferred by BLASTX searches and number of polymorphic sites found within the sequenced inbred lines are listed in Electronic Supplementary Material Online Resource 1. All primers were designed with Primer3 online software (Rozen and

Skaletsky 2000). Primers for CEL1CH were designed according to the specifications of Surveyor Mutation Kit™ (Transgenomics Inc., USA) and end-labeled with FAM (f) and/or Hex (h) (Online Resource 1). Fragments for dHPLC analysis were obtained using the primers formerly designed to sequence the control inbred lines (Fusari et al. 2008) which were in accordance with the specification of Oefner and Underhill (1998). Twelve candidate genes were further selected to assay the performance of the two methods by genotyping 22 sunflower accessions and 90 RILs.The length of the amplified regions, the haplotypes detected by Fusari et al. (2008) and the variable site positions for this set of candidate genes are listed in Table 1.

Genomic DNA extraction and polymerase chain reaction

DNA from 3-week-old plants grown in greenhouse was extracted from lyophilized leaves using Nucleon™ Phytopure™ genomic DNA extraction kit (GE Healthcare Life Sciences, Argentina) according to previously described protocols (Fusari et al. 2008).

PCRs were performed in a 50-µl volume reaction with 80–100 ng genomic DNA, 2 mM $MgCl_2$,

Table 1 Candidate genes selected for genotyping and mapping with CEL1CH and dHPLC

Method	Description	Length	SNPs and Indels	
	Blastx searches	**bp**	**Haplotypes Detected**[a]	**ID**
CEL1CH				
RS16	40S Ribosomal Protein S16	301	T A 206	1 2
1ACC0	1Aminocyclopropane 1-Carboxylic Acid Oxidase	450	T..G..T..C A..-..G..T 224 237 226 289	1 2
PGIP1	Polygaracturonase Inhibitor Protein	388	G..C..G C..T..A 78 345 318	1 2
MADSB-TF3	MADS-Box Transcription Factor	371	C G 152	1 2
GIA	Gibberellic Acid Insensitive-like Protein	569	T..C C..G 170 422	1 2
dHPLC				
CPSI	Chloroplast Photosystem I Reaction Center V	183	T..C C..G 107 150	1 2
CAB	Chloroplast Chlorophyll A/B Binding Protein	191	C..C T..T T..C 87 101	1 2 3
AALP	Arabidopsis Aleurain- like Protease	287	A..C..T..A..A..T..T..GGTTA..AATGTG..C..G..G G..C..C..G..G..T..C..GGTTA..AATGTG..C..G..G A..C..T..A..A..T..T..-----..AATGTG..T..G..G G..T..T..A..A..C..T..GGTTA..------..C..A..C 47 95 137 155 193-198 201 65 129 151 161-165 200 206	1 2 3 4
SCR1	Scarecrow Transcription Factor Type 1	812	G..A..C A..T..C A..A..A 108 751 561	1[b] 2 3
LIM	LIM Domain Containing Protein	324	G..T..-..C..A..A..G..---- G..G..-..T..G..T..A..---- A..T..T..C..G..T..A..TAAT 33 87 241 267 76 194 243 269-272	1 2 3
PGIC	Cytosolic Phosphoglucose Isomerase	204	A..A..A..T G..C..A..- A..C..G..- 109 151 117 169	1 2 3
CAM	Calmodulin	642	T..C..A..G..G..G..A..G..T..T..T..T..G..C..T..TTAC..C C..T..G..A..A..A..C..A..C..A..A..C..C..T..G..C..-----..A 21 64 140 148 165 277 361 437 467-480 58 136 141 161 212 354 413 453 560	1 2

[a] Numbers below nucleotides indicate the position of the polymorphism sites according to the alignment of all the haplotypes

[b] For *SCR1*, *LIM*, *PGIC* and *CAM*, the haplotype numbered 1 corresponds to the RHA266 parental line and the haplotype numbered 2 corresponds to the PAC2 parental line; the other haplotypes described were found within the sunflower inbred lines sequenced in a previous work (Fusari et al. 2008)

0.2 mM dNTP, 1 U *Taq* Platinum Polymerase (Invitrogen, Argentina) and 0.25 mM primer set. The cycling conditions were: 2 min at 94°C for initial denaturing, 35 cycles of 30 s at 94°C, 45 s at 65–58°C or 60–55°C, 1 min at 72°C and a final extension of 10 min at 72°C. The quality of extracted DNA and amplified products was visualized under ultraviolet light after electrophoresis on ethidium bromide-stained 1% and 2% agarose gels, respectively. In particular, the PCR products for treatment with CEL1 endonuclease were also scanned with a Typhoon Trio Scanner (GE Healthcare Bio-Sciences, Argentina) to ascertain the primer fluorescence labeling efficiency.

Heteroduplex formation

As previously mentioned, nine control inbred lines of known genotypes were used to test CEL1CH and dHPLC methods. Inbreds are supposed to be highly homozygous, and thus the heteroduplex molecules were made by mixing two PCR products of two inbred lines with different haplotypes for each locus. To characterize the 22 accessions of unknown genotype, heteroduplex samples were made by mixing each PCR product of the unknown inbred line with a PCR product of a control inbred line representative of each haplotype detected by Fusari et al (2008) (alleles A and B, Fig. 1). For the RIL population a similar procedure was carried out; every individual was mixed with RHA266 (parental line). Additionally, reference homoduplexes were made with PCR products from all samples including controls, accessions of unknown genotype, parental lines and RILs. All samples underwent the heating and re-annealing process by placing the tubes in a Mastercycler *ep*gradient S thermocycler (Eppendorf, Germany) running the following program: 95°C for 2 min, 95°C ramping to 85°C: –2°C/s, 85°C ramping to 25°C: –0.1°C/s, 4°C hold. The thermocycler has a cooling rate of –4.5°C/s, and thus the cooling ramp was set at 45% in the first step and at 2% in the second step. Samples were then either analyzed by dHPLC or incubated with CEL1 endonuclease followed by capillary electrophoresis to detect the corresponding polymorphisms. The unknown genotypes were inferred when the mixture unknown/control revealed a homoduplex-like profile, assigning to the line under assay the same genotype as the control line. For RIL analyses, heteroduplex formation in any mixture RIL/RHA266 was ascribed to the presence of the alternative haplotype derived from the PAC2 parental line, and was scored accordingly. When the mixture unknown/control inbred did not match any of the haplotypes tested, or when the homoduplex control of unknown inbred lines displayed an unexpected profile, the PCR products were sequenced as described later.

CEL1CH method

Partial purification of CEL1

CEL1 crude extract (CJE, celery juice extract) was obtained as previously described (Till et al. 2004, 2003). All steps were performed at 4°C. Chilled celery stalks (0.5 kg) were homogenized with a juice extractor. The juice was collected (total 0.36 l), adjusted to the composition of Buffer A (100 mM Tris-HCl, pH 7.7, 100 µM PMSF) and centrifuged at 2,600 *g* for 20 min. Solid ammonium sulphate ($(NH_4)_2SO_4$) was slowly added to the supernatant and gently stirred to a final concentration of 25% saturation. After 30 min, the suspension (0.4 l) was centrifuged at 16,000 *g* for 40 min at 4°C. The supernatant was adjusted to 80% saturation of $(NH_4)_2SO_4$. After 30 min of stirring, the mixture (total 0.44 l) was centrifuged at 16,000 *g* for 90 min. The pellets were re-suspended in Buffer A (0.1 starting volume). The suspension was transferred to a dialysis tube (Spectra/Por® 12–14,000 MWCO) and dialyzed against a total of 32 l of the same buffer with four changes over 4 h. Aliquots of CJE (0.5 ml) were stored at –20°C. These aliquots conserved the activity despite the freeze–thaw process that is a consequence of the method.

CEL1 cleavage of heteroduplex

After PCR amplification and heteroduplex formation, different amounts of DNA FAM-G and FAM-C control mixtures (100–500 ng of total DNA) were tested for digestion with different volumes of CJE (0.2 µl, 0.5 µl, 1 µl and 2 µl, Fig. 2). The samples were assayed in a total volume of 20 µl supplemented with 2 µl of 10 × CJE reaction buffer (5 ml of 1 M $MgSO_4$, 5 ml of 1 M 4-(2-hydroxyethyl)-1-piperazine-ethanesulfonic acid (HEPES), pH 7.5, 2.5 ml of 2 M KCl, 0.1 ml of 10% Triton® X-100,

a Optimization of CJE volume

CJE	DNA	SNP cleavage of FAMC-FAMG		
		Heteroduplex		Homoduplex
µL	ng	5' 217bp	5' 415 bp	632 bp
	100	++	++	++
0.2	200	++	-	+
	500	+++	+++	+++
	100	+	-	-
0.5	200	+	-	+
	500	+++	++	+++
	100	-	-	-
1	200	+	-	-
	500	++	+	++
	100	-	-	-
2	200	-	-	-
	500	+	-	-
	100	-	-	-
5	200	+	-	-
	500	-	-	-

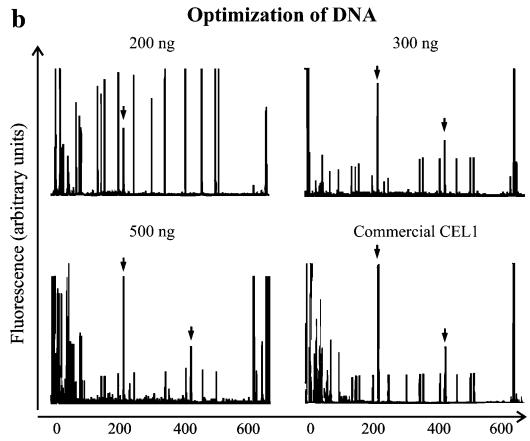

Fig. 2 Optimization of heteroduplex cleavage with in-house CEL1. **a** In-house CJE activity was tested by incubating 0.2–5 µl of enzyme with increasing amounts of the PCR amplicon mixture of FAM-G and FAM-C control DNAs (100–50 ng) (Transgenomics Inc.). The detection of all cleaved fragments and the homoduplex molecules were scored as minus (no detection) or plus (detection). **b** CJE performance (0.2 µl) with 200, 300, 500 ng of mixed amplicons is shown together with the digestion profile obtained with commercial CEL1 (1 µl of Surveyor® Nuclease and 200 ng of mixed amplicons). *Arrows* indicate the fragments expected at 217, 415 bp (cleavage of heteroduplex molecules) and 632 bp (homoduplex molecules)

5 µl of 20 mg/ml bovine serum albumin, 37.5 ml deionized water). The samples were incubated in a heating block at 45°C for 15 min and stopped with 5 µl of stop solution (0.15 M ethylenediaminetetraacetic acid (EDTA), pH 8.0). The product was then precipitated with 2.5 volumes of absolute ethanol at –20°C for 30 min, centrifuged at 3,600 *g* for 45 min and re-suspended in 5 µl of deionized water. The FAM-G and FAM-C controls are PCR FAM-labeled fragments of 632 bp; each one has a G or C at 217 bp from its 5′ end. The activity of CJE was confirmed when three peaks were detected at 217 bp, 415 bp and 632 bp (homoduplex length). A similar procedure was carried out for candidate regions.

Capillary electrophoresis

The final CJE digestion product (5 µl) was diluted with 10 µl of Hi-Di formamide (Applied Biosystems, USA) and 0.25 µl of GeneScan 500 (-250) ROX size standard (Applied Biosystems, USA). Samples were heated at 95°C for 5 min and cooled on ice for 5 min. Each mixture was injected into the ABI 3130xl Genetic Analyzer (Applied Biosystems, USA) and was separated across a 50-cm capillary array containing POP-7 polymer (Applied Biosystems, USA). The results were analyzed using GeneMapper application software v3.7 (Applied Biosystems, USA).

dHPLC analysis

The dHPLC analysis was performed using an Agilent series 1100 HPLC system (Agilent Technologies Inc., USA) equipped with a biocompatibility kit, a solvent degasser unit, a binary pump, an autosampler with cooling module (set to 4°C during analysis), a column oven and variable wavelength detector. Homo- and heteroduplex separation was performed using a Varian Helix™ DNA column (3 × 50 mm, Varian Argentina Ltd., Argentina). Chromatography was assayed at flow rates of 0.45 and 0.9 ml min^{-1}, with a linear binary gradient of Varian Helix BufferPaks A and B (Varian Argentina). DNA duplexes were detected at a wavelength of 260 nm. Partial denaturing temperatures for SNP scanning were determined based on the sequence of control inbred lines using dHPLC Melt Program available at the Stanford University website (http://www.insertion.stanford.edu/melt.html). The optimal temperature to detect variations for each gene was determined empirically by testing the heteroduplex samples obtained by mixing two control inbred lines with different haplotypes. The dHPLC system was tested before each series of runs by injecting a puC18 HaeIII digest (Varian Argentina) and DYS271 SNP standards (Varian Argentina) which were analyzed at 50°C and 56°C following the supplier's recommendations. Peaks were detected using the following integration

parameters: Slope sensitivity 1; Peak width 0.04; Area reject 1; Height reject 0.7; Shoulders off.

Sequencing

PCR products were purified using QIAquick PCR Purification Kit (Qiagen, Germany) and both strands were sequenced using specific primers on an ABI 3130xl Genetic Analyzer (Applied Biosystems, USA). ABI trace files were aligned using ABI Prism SeqScape Software v2.5 (Applied Biosystems, USA). SeqScape quality values of base-calls were set $\geq$20, the identification of mixed bases (heterozygous positions) was set at 25% or higher for second peak height, and default settings for the remaining parameters were used for SNPs and indel discovery. Heterozygous positions were re-checked with chromatogram files.

Genetic map construction

The candidate genes genotyped in the RIL population by dHPLC were added to the available genetic map derived from the cross PAC2 × RHA266 (Poormohammad Kiani et al. 2007a), using Mapmaker 3.0 (Lander et al. 1987) and CarthaGene 0.999 software (Schiex and Gaspin 1997). Chi-square tests were performed for segregation distortion analysis of each locus. Loci were assembled into groups using likelihood odds (LOD) ratios with a LOD threshold of 4.0 and a maximum recombination frequency threshold of 0.35 (Poormohammad Kiani et al. 2007a). The Kosambi mapping function was used to calculate map distances (cM) from recombination frequencies (Kosambi 1944). Estimated recombination frequencies of RILs were corrected with the Haldane function (Haldane and Waddington 1931). Mapchart 2.1 was used for graphical presentation of linkage groups and map position of functional markers (Voorrips 2002).

Results

To test the performance of CEL1CH and dHPLC as genotyping methods in sunflower, we evaluated 107 SNPs and 23 indels corresponding to 24 candidate genomic regions previously sequenced in 19 sunflower inbred lines (Fusari et al. 2008) (Online Resource 1).

CEL1CH method

The CEL1 endonuclease used in this work was prepared in-house from celery stalks following a two-day protocol as described in the Materials and Methods section. The amount of total DNA needed to detect cleavage using CJE was assayed by incubating 100, 200 and 500 ng of a mixture of FAM-C and FAM-G PCR products with different volumes of the enzyme extract (0.2–5 µl, Fig. 2). The expected digestion patterns were obtained for enzyme volumes of 0.2, 0.5 and 1 µl and 500 ng of total DNA. As expected, a weaker signal but a correct pattern of cleavage was seen with 0.2 µl of CJE and 100 ng of total DNA. Higher volumes of CJE (2 and 5 µl) showed either no peaks or unspecific digestion (Fig. 2a). In order to adjust the DNA quantity, further assays were carried out using 0.2 µl of CJE in reactions containing 300, 350, 400, 450 and 500 ng of total DNA (Fig. 2b). The electropherograms showed that 0.2 µl CJE cleaves 300 and 500 ng of total DNA as efficiently as the commercial CEL1, 300 ng being the smallest amount of DNA with the highest signal/background ratio (Fig. 2b). Therefore, the latter conditions were used to assay CEL1CH performance.

A total of 13 loci corresponding to genes *LIM*, *MADB-TF3*, *PGIP1*, *PGIP3*, *CAB*, *GIA*, *GPX*, *PSI-III-CAB*, *SCR1*, *RS16*, *LZP*, *GO* and *1-ACCO* were suitable for genotyping using CEL1CH according to the criteria established by the Surveyor Mutation Kit™ supplier (Transgenomics Inc., USA) (fragment size $\geq$ 200 and $\leq$ 800 bp, a minimum distance of 50 bp between the primers and the adjacent SNP and distances in-between SNPs). The remaining sequences were not evaluated further by this approach because they did not satisfy these technical requirements regarding fragment size (*CPSI*, *PGIC*), distance between primers and the adjacent SNPs (*SEM*, *GcVT*, *AALP*, *SCR2*, *ANT* and *LHCP*), number of SNPs in the region (*NsLTP*, *CAM*) and distance in-between SNPs (*GAPDH*) (Online Resource 1). Initial tests were performed on a set of nine control inbred lines of known genotype with oligonucleotide primers that were 5′ end-labeled with FAM or HEX. All primer pairs produced an abundant single PCR product, except for the loci *MADSB-TF3*, *PGIP1*, *PSI-III-CAB* and *CAB*. A second PCR round was carried out using only one primer 5′ end-labeled primer for these

four regions. *MADSB-TF3*, *PGIP1* and *PSI-III-CAB* amplicons were obtained in high yield; however, *CAB* amplification was scarce, hence, it was excluded from subsequent analysis with CEL1CH.

Thus, CEL1CH methodology was succesfully employed to detect polymorphism in 10 out of the 12 tested candidate regions, with all control lines being properly genotyped by this method. These regions are good representatives of different polymorphism situations since they have a single SNP (*MADSB-TF3* and *PSI-III-CAB*), two SNPs (*PGIP3*, *RS16*, *GIA* and *SCR1*), 8-bp indel (*ZFP*) or complex polymorphism patterns (*PGIP1*, *1-ACCO*, *LIM*) (Fig. 3a and Online Resource 3). Interestingly, although the *GO* region was cleaved correctly by CJE, differences in length among the homoduplex molecules themselves allowed genotyping by capillary electrophoresis alone, making the cleavage step unnecessary. On the other hand, for the *GPX* region, CEL1CH failed to detect an indel polymorphism of 6 bp.

The length of the fragments identified after capillary electrophoresis was in agreement with the position of the mismatch caused by a given SNP or indel. Homoduplex controls only showed the peak which corresponded to the full-length fragment (Table 1; Fig. 3a, Online Resources 1 and 3). Although in most of the cases both the forward and reverse primers were individually labeled with different fluorophores, in

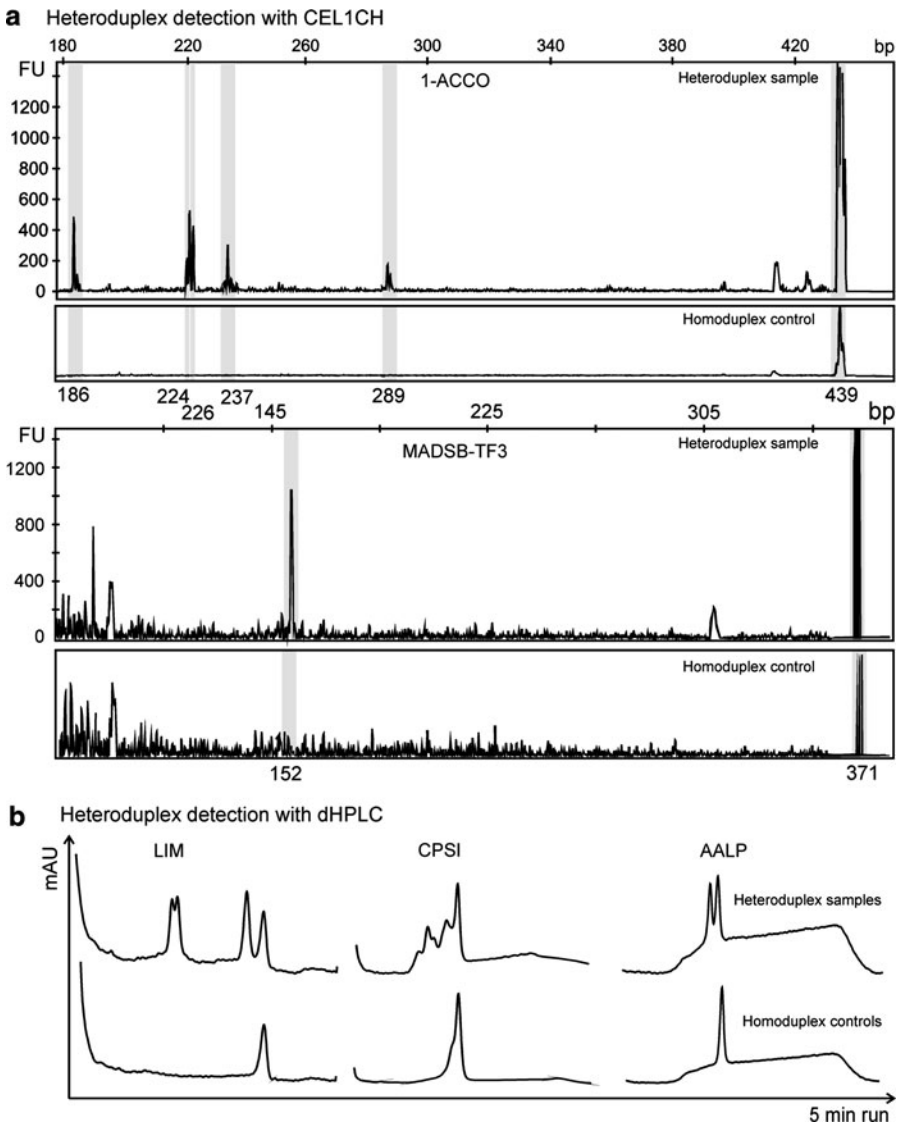

Fig. 3 Detection of sunflower SNPs with CEL1CH and dHPLC. **a** CEL1CH genotyping method for two candidate genes (*1-ACCO* and *MADSB-TF3*). For each gene, the upper panel shows analysis of heteroduplex samples and the bottom panel shows homoduplex controls. *Grey bars* indicate the cleavage or homoduplex fragment at the corresponding base-pair position. The *y* axes are in Fluorescence Units and *x* axes are in base pairs. **b** dHPLC genotyping method for three candidate genes (*LIM*, *CPSI* and *AALP*). For each gene, the upper chromatogram corresponds to elution profiles of heteroduplex samples and the bottom chromatogram to the homoduplex controls both obtained at 0.9 ml min^{-1} in 5 min of running (*x* axes). *y* axes are in milli-Absorbance Units (mAU) that correlate with milli-Volt Units

some instances just one labeled primer was sufficient to unambiguously detect the polymorphisms.

Due to their distinctive patterns of polymorphism and the superior performance of the CEL1CH method for candidate genes *RS16*, *1-ACCO*, *PGIP1*, *MADSB-TF3* and *GIA* relative to the remaining genes studied, these regions were further selected to genotype a pool of 22 sunflower accessions of unknown genotype, to prove the scalability of the methodology.

The gene IDs, amplicon lengths, variable positions and haplotypes detected by Fusari et al. (2008) are given in Table 1.

dHPLC method

SNP genotyping by dHPLC was evaluated by using 20 out of the 24 candidate genes (Online Resource 1). The remaining four candidate genes had previously been discarded for different reasons. Two of them were already efficiently genotyped by CEL1CH (*RS16* and *LZP*); *1-ACCO* did not fulfill the method requirements, as the running conditions could not be predicted due to the incomplete sequence in control lines. Finally, *GO* can be genotyped by other methods simpler than heteroduplex analysis. High PCR product yields were obtained with all primer pairs tested except for one pair designed for *PGIP1* amplification (511-bp fragment); hence, the *PGIP1* gene was only tested with a 301-bp fragment (Online Resource 1). Runs were performed on a standard HPLC system. Two flow rates were assayed to determine the optimal run time. The first series of experiments was conducted at a flow rate of 0.45 ml min^{-1}. These runs lasted 11 minutes and the fragments began to elute at 6–7 minutes (data not shown). Afterwards, the flow rate was increased to 0.9 ml min^{-1}, leading to a faster elution of the molecules without losing resolution; consequently, the run time was reduced from 11 to 5 minutes, including signal re-balance time. The dHPLC protocol optimization also involved the adjustment of sample DNA concentration and injected volume. These analyses determined that the minimal amount of DNA to inject was 100 ng, in an injection volume of 5–10 µl (about 20 ng μl^{-1}). Optimal temperature and gradient differed according to the base composition of the sequence under study, with most of the genes falling within the range of 54–61°C (Online Resource 2). The parameter combinations tested allowed discrimination of homo- and heteroduplex molecules for the genes *CPSI*, *GcVT*, *AALP*, *SCR2*, *LIM*, *MADSB-TF3*, *CAB*, *SCR1*, *LHCP*, *CAM* and *PGIC* (Online Resource 2). For the remaining nine genes it was not possible to distinguish between homo- and heteroduplex molecules under any of the assayed conditions (*SEM*, *PSI-III-CAB*, *GIA*, *GPX*, *PGIP1*, *PGIP3*, *ANT*, *NsLTP* and *GAPDH*). For *SCR1* and *CAM*, two combinations of temperature and buffer B gradient were needed to reveal all the haplotypes found within the genotypes included in the analysis (Online Resource 2). Regarding heteroduplex elution profiles, *LIM* and *CPSI* showed the expected pattern of four peaks (two heteroduplex molecules and two homoduplex molecules, Figs. 1 and 3b). *SCR1* and *GcVT* heteroduplex molecules eluted as three peaks. For the remaining regions, heteroduplex profiles showed only two peaks (*AALP*, *MADSB-TF3*, *CAB*, *PGIC*, *LHCP*, *SCR2* and *CAM*, Fig. 3b and Online Resource 4). According to the fundamentals of the technique, the peak that elutes earlier from the column corresponds to the heteroduplex molecules, while the peak that elutes later in the column corresponds to the homoduplex molecules. In fact, homoduplex profiles displayed only one peak, with a retention time corresponding to the peak that eluted later in the heteroduplex samples. Particularly for *SCR2* and *MADSB-TF3*, the homoduplex samples displayed an unexpected profile since each of them showed an extra peak. The extra-peak elution times were earlier than heteroduplex elution times; therefore they were not interpreted as heteroduplex molecules (Online Resource 4).

From the pool of genes optimized for detection with dHPLC, seven were selected for subsequent scale-up analysis (*CPSI*, *CAB*, *AALP*, *SCR1*, *LIM*, *PGIC* and *CAM*). Thus, six loci were used to genotype a set of 22 accessions with unknown genotype, and four of them that had previously been shown to reveal polymorphism between RHA266 and PAC2 parentals were also genotyped in 90 RILs to map functional markers (*SCR1*, *LIM*, *PGIC* and *CAM*). The corresponding gene IDs, amplicon lengths, variable positions and the haplotypes detected by Fusari et al. (2008) are given in Table 1.

Genotyping of germplasm accessions with CEL1CH and dHPLC

CEL1CH was used to genotype *RS16*, *1-ACCO*, *MADSB-TF3*, *PGIP1* and *GIA* regions, while *CPSI*, *CAB*, *LIM*, *SCR1* and *PGIC* were genotyped with dHPLC method.

The genotyping results obtained for the 22 sunflower inbred lines are described in Online Resource 5. A total of 110 genotypes were determined by the CEL1CH method (5 loci per 22 inbred lines), whereas 131 genotypes were obtained by dHPLC analysis, failing only to genotype the *LIM* locus for the inbred HAR4 (Online Resource 5).

A detailed inspection of Online Resource 5 shows that the 22 inbred lines possess only one variable site at position 206 for the *RS16* locus, while two polymorphic sites were detected in the control inbred lines (Fusari et al. 2008) (Online Resource 1). Similarly, despite regions *PGIP1* and *1-ACCO* having four and five potential polymorphic sites, polymorphisms were detected only for three and four sites, respectively, for the group of lines studied here (Online Resource 1 and 5). In addition, *RS16*, *MADSB-TF3*, *SCR1* and *PGIC* were heterozygous for the inbred lines HA336, HAR4, R432 and HA853 (Online Resource 5). The heterozygous loci were identified by both CEL1CH and dHPLC by analyzing the supposedly homoduplex profiles. As expected, heterozygosity led to the formation of heteroduplexes and they were either cut by CJE in CEL1CH or displayed a heteroduplex-like profile on dHPLC analysis. All heterozygous genotypes were confirmed by direct sequencing. For MP557, P94, F164B and HA64 inbred lines, the homoduplex controls behaved properly, but their genotypes could not be determined. Hence, they were analyzed by direct sequencing, revealing the presence of new haplotypes. These additional SNPs and/or indels provide new data on *CAB*, *LIM* and *SCR1* genes (Online Resource 5).

SNP mapping with dHPLC

Four candidate genes analyzed with dHPLC (*SCR1*, *LIM*, *PGIC* and *CAM*) were used to genotype 90 RILs from a mapping population (Poormohammad Kiani et al. 2007a). The four loci were assigned to different linkage groups (LG) as shown in Fig. 4. *SCR1* was included in LG6, at 3.4 and 1.3 cM from the adjacent markers ORS381 and ORS1233, respectively. *PGIC* was located at the distal end of LG9 at 5.8 cM from the marker ORS844. *LIM* was inserted in LG10, and the distances between it and the adjacent markers WHA3204 and HA2579 were 1.2 and 2.6 cM, respectively. Finally, *CAM* was anchored in LG13, between two distant markers E35M49_10 and HA4208, at 6.2 and 7.5 cM from them, respectively.

Discussion

Heteroduplex analysis based on both CEL1CH and dHPLC has been proven to permit low-budget academic and molecular breeeding laboratories to improve the throughput of SNP genotyping in plants and animals at reasonable cost. Despite the fact that these methods have been used widely in plant reverse genetic approaches, there are no records of SNP genotyping by CEL1CH in sunflower and only two studies have used dHPLC to genotype SNPs or indels in this crop (Lai et al. 2005; Lexer et al. 2003). Meanwhile, the number of functional markers (SNPs and indels) that could potentially be exploited for germplasm characterization and association mapping studies continues to grow.

Theoretically, given a candidate gene, several regions within it can be targeted for analysis either by CEL1CH or dHPLC. In practice, however, the choice will be ultimately determined by the requirements, potentialities and drawbacks of each technique. Generally, the issues to be taken into consideration include the number of variable sites within the region, the amplicon length, the knowledge of the complete sequence and the economic costs of each method.

The possibility of producing in-house CEL1 endonuclease greatly reduces the cost of the CEL1CH method without loss of sensitivity. As reported by Galeano et al. (2009), celery stalks are a good source of the enzyme because they have a low level of interfering pigments. The CJE obtained was used to cleave mismatches in heteroduplex molecules and its activity was as efficient as the commercial CEL1 (Fig. 2). The effectiveness of CJE agrees with the findings of Oleykowski et al. (1998), who showed that CEL1 reliably cleaves mismatches, cutting DNA at the 3′ side of single base-pair substitutions at DNA distortions. Although CEL1 also contains 5′ to 3′ exonucleolytic activity and could digest the full-

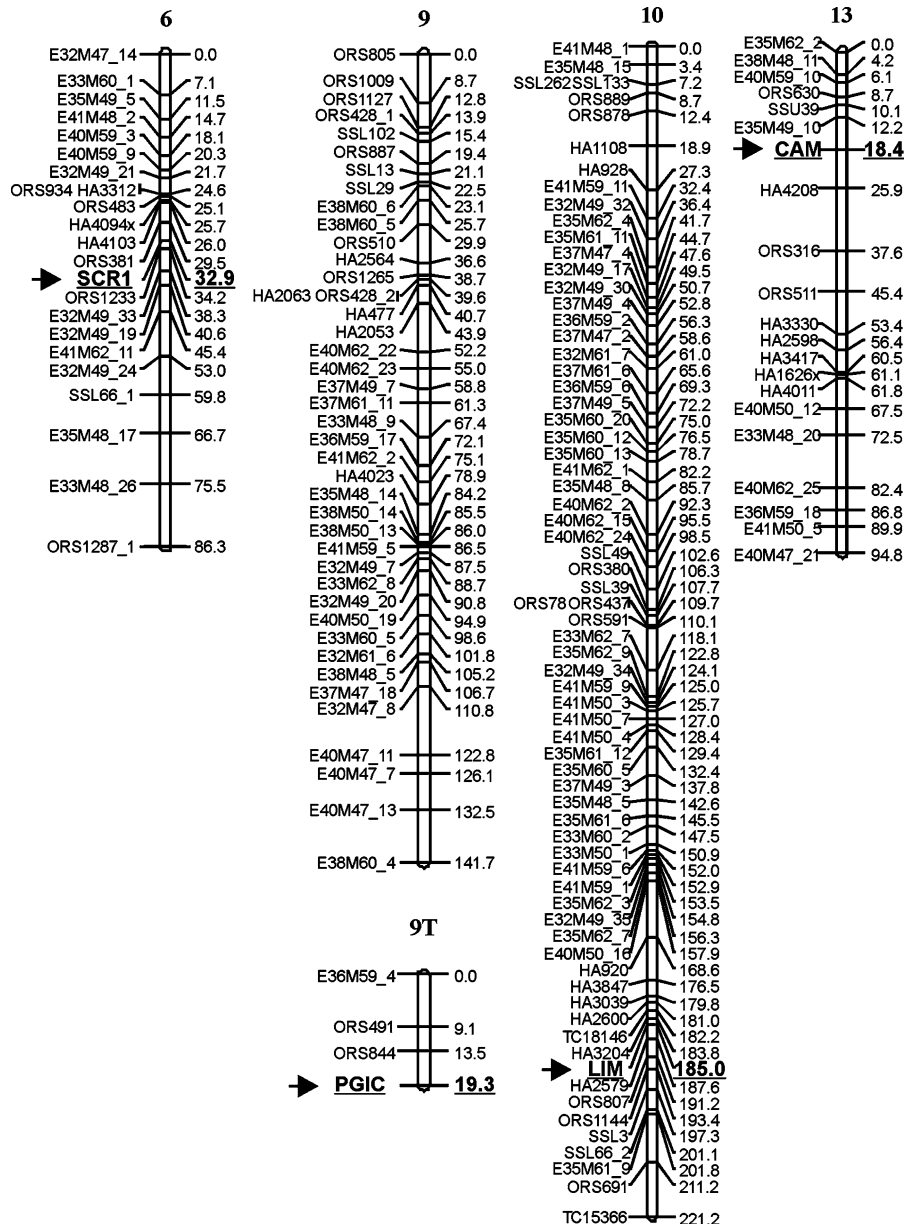

Fig. 4 Candidate genes anchored in sunflower linkage map. Linkage groups 6, 9, 10 and 13 of sunflower map PAC2 × RHA266. The selected candidate genes *SCR1*, *PGIC*, *LIM* and *CAM* are shown underlined. LG10 anchors two markers (TC18146, TC15366) corresponding to sequences of the TIGR Gene Indices Tentative Consensus v. 2010 not included in the map of Poormohammad Kiani et al. (2007b)

length PCR product starting with the 5′ fluorescent label (Barkley and Wang 2008), the incubation time set in this work (15 min) prevented CEL1 exonuclease activity and the loss of the fluorescent signal (Fig. 2). One of the initial large investments for CEL1CH experiments is the costly instruments commonly used for cleaved-fragment detection, such as the LI-COR DNA analyzer. In our case, the experiments were performed using a capillary electrophoresis platform previously used for SSR genotyping (Genomic Unit, INTA). Despite the fact that labeled primers are needed, we were able to detect SNPs and indels with only one labeled primer, thus reducing the costs of this step. The capillary electrophoresis running cost was calculated at USD 2.5 per data point. This relatively high cost can be reduced by marker multiplex loading, combining different fluorescent labels and different patterns of cleavage depending on the molecular characteristics of the regions under study.

Despite the huge impact of reverse genetics approaches (Cooper et al. 2008; Perry et al. 2003;

Till et al. 2007; Uauy et al. 2009; Wang et al. 2008; Xin et al. 2008), CEL1CH has been only rarely used in the evaluation of diversity in natural populations, probably due to the restrictions regarding amplicon length, number of SNPs present per region, and demanding specifications of primer design. Nevertheless, the successful assessment of diversity in *Arabidopsis* ecotypes over five genes (Comai et al. 2004) has allowed the technique to expand to a few different plant species such as melon, poplar, common bean and tomato (Galeano et al. 2009; Gilchrist et al. 2006; Nieto et al. 2007; Yang et al. 2004). On the other hand, for rice, SNP identification efficiency by mismatch cleavage of heteroduplex was lower than the effectiveness reached by PCR-RF-SSCP (PCR-restriction fragment-single strand conformation polymorphism), although Shirasawa et al. (2007) used another source of endonuclease. In this work, out of the 24 regions evaluated for CEL1CH genotyping, 10 were efficiently optimized for the detection of cleavage fragments derived from mismatch heteroduplex molecules (Fig. 3a and Online Resource 3). After adjusting the technique requirements, the performance of the CEL1CH method was satisfactory, allowing the identification of polymorphisms ranging from one variable site in the *MADSB-TF3* gene to more complex patterns of polymorphism for *1-ACCO*, *PGIP1* and *LIM* (Fig. 3a and Online Resource 3). However, some regions could not be subjected to analysis by this method because either a short distance between SNPs (*GAPDH*) or a high number of SNPs in the region (*NsLTP*, *CAM*) impeded recognition of all the fragments derived from CEL1 cleavage. Even though the use of CEL1 to date is generally restricted to detecting single mutations over a gene of interest (Cooper et al. 2008; Nieto et al. 2007; Uauy et al. 2009), we were able to simultaneously genotype up to five SNPs in a single amplicon including 2- to 8-bp indels.

Barkley and Wang (2008) suggested that the limitations imposed by amplicon size and the difficulty of tracking SNPs at the ends of the target amplicons could be overcome by designing overlapping regions to ensure the detection of all sites in a target gene. In this work, some regions were discarded for analysis with the CEL1CH method after all possibilities for primer design or overlapping fragments were examined. After analyzing the features of the regions under study, two were considered unsuitable for genotyping with CEL1CH due to the amplicon size (*CPSI* and *PGIC* $\leq$ 200 bp), and five were discarded because the distance between primers and the adjacent SNPs was less than the minimum required (*SEM*, *GcVT*, *AALP*, *SCR2*, *ANT* and *LHCP*). On the contrary, two candidate genes over 1,000 bp in length were amplified and genotyped over an inner region of ca. 371 and 488 bp without difficulties (*MADSB-TF3*, *PGIP1*).

The traditional CEL1CH method involves the end-labeling of forward and reverse primers, because the cleavage analysis of the two strands serves as a confirmation and control within the same experiment. However, amplifications carried out with only one end-labeled primer did not compromise the detection of cleavage fragments for *MADSB-TF3*, *PGIP1* and *PSI-III-CAB*.

The main advantage of the CEL1CH method is that it can determine the specific location of SNPs without using any toxic chemicals. CEL1CH seems to be an exceptionally powerful technique when it is important to trace the precise location of the variable site rather than identify haplotypes, such as reported by Hung et al. (2008) in the detection of point mutations involved in beta-thalassemia disease. Although the amount of total DNA needed to perform the experiments has been a drawback compared with other techniques, a recent report revealed the optimization of TILLING and EcoTILLING in potato with DNA amounts ranging from 25 to 0.5 ng (Elias et al. 2009).

As mentioned previously, CEL1CH efficiently genotyped the 10 tested loci and, moreover, it was sensitive enough to detect heterozygous individuals for genes *RS16* and *MADSB-TF3* (Online Resource 5). Interestingly, *PGIP1* and *1-ACCO* sequences included four and five polymorphisms in the control inbred lines, but three and four SNPs, respectively, were detected when assaying the group of 22 unknown-genotype accessions. Although CJE was not 100% efficient in these two cases, information about variable sites was obtained anyway, since cleaved fragments were independently detected (Online Resource 3 and 5). Unlike the dHPLC technique, CEL1CH succeeded in genotyping both *PGIP1* and *1-ACCO*. The dHPLC method failed to detect differences between heteroduplex and homoduplex for the *PGIP1* gene and to predict running conditions for *1-ACCO*, since sequence data was incomplete.

The advantages of the dHPLC technique are independence from labeled primers, adaptability to semi-automated systems, high reproducibility and sensitivity and the fact that it provides an objective criterion for sample classification (Abbas et al. 2004). A number of studies reported its effectiveness and specificity in comparison with PCR-RFLP (PCR-restriction fragment length polymorphism), direct sequencing and *s*ingle strand conformation polymorphism (SSCP) (Abbas et al. 2004; Jones et al. 1999; Yu et al. 2005). In addition, Costabile et al. (2006) reviewed different techniques used in the diagnosis of primary immunodeficiency diseases, emphasizing the greatest accuracy of the dHPLC method. The dHPLC method is able to identify haplotypes rather than individual SNP alleles, which in some instances, such as genotype–phenotype association analysis, can be more informative than single-marker data (Rafalski 2002). In this work, 11 out of 24 candidate gene regions were optimized for heteroduplex detection with dHPLC and seven were further selected to genotype germplasm accessions or to add functional markers to a sunflower linkage map (Online Resource 2, Fig. 4). It is remarkable that for most of the regions, one temperature-gradient condition was sufficient to detect all differences between the haplotypes tested, except for *SCR1* and *CAM*. While differences between *SCR1* haplotype pairs 1–2 and 1–3 were detected with the temperature set at 59°C; the mixture of *SCR1* haplotypes 2–3 required the temperature to be set at 61°C, adjusting the buffer B gradient accordingly (Table 1 and Online Resource 2). Similarly, two out of the seven *CAM* haplotypes identified by Fusari et al. (2008) were distinguished from each other only when the temperature was increased to 60°C (Online Resource 2). This additional setting was probably necessary because the only difference between these two haplotypes was an 8-bp indel (Fusari et al. 2008). Despite the fact that fragment length is one limiting condition for dHPLC, other considerations must be taken into account to succeed with this technique. The complete and precise sequence of the region under analysis must be known to properly predict the partial denaturing condition, a constraint which does not apply to the CEL1CH method, since in this case the critical factors are the length and SNP positions in the region. Although dHPLC is more versatile than CEL1CH for identifying multiple SNPs per region, the number of potential haplotypes produced by a given number of SNPs is an additional limitation for this technique. For instance, seven distinct haplotypes were found for *CAM* (combining 20 SNPs and six indels), and thus genotyping this region entails heteroduplex analysis of seven different mixtures for each unknown inbred line. Notwithstanding the difficulties of genotyping *CAM* in the 22 unknown accessions, it was unambiguously genotyped in the 90 RILs from the PAC2 × RHA266 mapping population, given that only the two haplotypes corresponding to the parental lines were present (17 SNPs and one indel, Table 1). Nevertheless, this number of SNPs is still prohibitive for genotyping *CAM* with the CEL1CH method. The dHPLC method allowed the identification of new haplotypes when none of the mixtures unknown-line/control-line displayed a homoduplex-like profile, while homoduplex controls behaved properly. Direct sequencing confirmed that in regions like *CAB* and *SCR1* new polymorphic sites were present, whereas in *LIM* the new haplotype detected had a different combination of the polymorphisms already identified in the sequences of control inbred lines (Online Resource 5). Although dHPLC has been usually applied to genotype SNPs already identified in candidate genes, the fact that new haplotypes were found here offers a promising perspective for de novo SNP identification (Han et al. 2004).

The dHPLC method was also able to detect heterozygous individuals (*SCR1* and *PGIC*, Online Resource 5). As previously mentioned, both CEL1CH and dHPLC were effective in detecting heterozygosity in the heavily inbred lines analyzed here. Besides aiding in the discovery of hybridization or introgression events, this feature is particularly attractive for functional genomics analyses since heterozygosis likely defines functionally important regions of relevant genes (Weil and Monde 2007).

Estimated running costs of dHPLC reach USD 0.8 per data point. This is remarkably low compared to any of the high-throughput methods available to identify SNPs to date (Bagge and Lübberstedt 2008), which makes it ideal for bi-parental population mapping. In fact, Kota et al. (2001) tested dHPLC mapping performance with four ESTs (expressed sequence tags) previously mapped as RFLP markers, obtaining the same positions in the barley map; and afterwards they mapped 216 EST-derived SNPs and four indels in three mapping populations (Kota et al.

2008). In sunflower, dHPLC has been used to map 11 genes and 243 ESTs (Lai et al. 2005; Lexer et al. 2003), endorsing the feasibility of the technique. In this study, four markers that were polymorphic in the parental lines were genotyped and unambiguously located in different LGs. In fact, the candidate gene *CAM* was located in LG13, filling a 13 cM gap. In addition, *PGIC* was anchored with a small group of markers located at the distal region of LG9 in the map obtained by Tang et al. (2003) and with a sub-group of markers in LG9 obtained by Poormohammad Kiani et al. (2007a) (Fig. 4). Candidate gene *LIM* was located in LG10, where six QTLs explaining 13% of the sunflower resistance to *S. sclerotiorum* fungal pathogen were found (Maringolo, Master Thesis 2007). This gene is a putative transcription factor that could be involved in defense mechanism, although it was also described as been involved in the formation of cytoskeleton (Arnaud et al. 2007). Despite the fact that the sunflower map of PAC2 × RHA266 still needs saturation for some LGs, the correlation with other Compositae maps can be done based on the public markers (mainly SSRs) flanking the functional markers described in this work, through the tools available at the Compositae Genome Project web site (http://www.sunflower.uga.edu/cgi-bin/cmap/feature_search).

In summary, CEL1CH and dHPLC are complementary methods for genotyping SNPs and indels in sunflower, since three candidate regions were genotyped by both techniques while the others were uniquely genotyped by either dHPLC or CEL1CH (seven genes each method). For instance, dHPLC complements CEL1CH in the analysis of those regions with multiple polymorphisms where CEL1CH loses efficiency (e.g., *CAM*, *LIM* and *AALP*). On the other hand, sites with fewer SNPs or indels are easily optimized for CEL1CH (such as occurred with *GIA*, *MBTF*, *PGIP3* and *RS16*). However, not all the regions with complex haplotypes could be analyzed using dHPLC (e.g., *PGIP1*, *NsLTP*). Therefore, CEL1CH serves as a good alternative strategy for genotyping at least one of the polymorphic positions present in these regions. Neither CEL1CH nor dHPLC required large investments in terms of consumables, since CEL1 is easily obtained in the laboratory and dHPLC only requires PCR amplification and fragment mixture. The investment in equipment is not a determining issue for CEL1CH, since it can be outsourced from a sequencing service; other less expensive but more time-consuming procedures, such as agarose or acrylamide gel electrophoresis, can be applied to detect the presence of cleaved fragments. Therefore, CEL1CH provides a rapid and inexpensive way to detect natural diversity in candidate genes. The optimization process in dHPLC necessitates assaying at least two to four temperature-gradient conditions surrounding the parameters estimated by the software, making it difficult to implement the technique without laboratory equipment. The cost of purchasing specific dHPLC equipment may be a factor preventing the wide application of this strategy; however, to circumvent this problem, we used an analytical HPLC machine and a specific DNA column as described by Randall et al. (2005), obtaining a very good performance.

Since marker genotyping methods were first developed in the 1980s, many techniques have been refined and optimized depending on budget, equipment, and personnel (Collard et al. 2008). In this context, CEL1CH and dHPLC become two complementary techniques that must be carefully chosen depending on the selected candidate genes and the polymorphisms being genotyped. Nonetheless, they are both good and cost-effective SNP genotyping alternatives for molecular breeding labs to improve the number of markers and individuals to be analyzed in a medium- to high-throughput manner. Moreover, where possible, setting up both techniques in the laboratory would allow genotyping of almost 100% of the candidate genes of interest, as described above.

The results presented here indicate that CEL1CH and dHPLC are cost-effective, broadly applicable, scalable and efficient methodologies for exploiting functional markers in sunflower. Both methods demonstrated their suitability for exploratory approaches in the first evaluation of variability for a candidate gene set and also, as reported here, as methods for expanding polymorphism genotyping of candidate genes over both a wider group of germplasm and a bi-parental population. Finally, this is the first report of EcoTILLING-like studies in sunflower, providing good evidence of its potentiality for candidate gene diversity characterization. Moreover, the availability of these tools, together with other SNP genotyping techniques, could expand the TILLING approach recently applied in discovering new imidazolinone-resistant sunflower (Sala et al. 2008) to other agronomically important traits in this crop.

Acknowledgments This research was supported by ANPCyT/FONCYT, PID 2007 00073, INTA-PRR AEBIO 245001 and 245005, INTA-PE AEBIO 24554711 and 241351. The authors thank the Germplasm Bank from Estación Experimental Agropecuaria INTA Manfredi for kindly providing sunflower inbred lines seeds. C.M.F. is a Ph.D. student supported by a fellowship from Instituto Nacional de Tecnología Agropecuaria (INTA). Drs. V.V.L., R.A.H. and N.B.P. are career members of the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Dr. H.E.H. is a career member of the Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC) and Professor at the Facultad de Ciencias Exactas y Naturales, University of Buenos Aires (UBA).

References

Abbas A, Lepelley M, Lechevrel M, Sichel F (2004) Assessment of DHPLC usefulness in the genotyping of GSTP1 exon 5 SNP: comparison to the PCR-RFLP method. J Biochem Biophys Meth 59:121–126

Al-Chaarani GR, Gentzbittel L, Huang XQ, Sarrafi A (2004) Genotypic variation and identification of QTLs for agronomic traits. using AFLP and SSR markers in RILs of sunflower (Helianthus annuus L.). Theor Appl Genet 109:1353–1360

Al-Chaarani GR, Roustaee A, Gentzbittel L, Mokrani G, Barrault G et al (2002) A QTL analysis of sunflower partial resistance to downy mildew (*Plasmopara halstedii*) and black stem (*Phoma macdonaldii*) by the use of recombinant inbred lines (RILs). Theor Appl Genet 104:490–496

Alfadil AT, Poormohammad Kiani G, Dechamp-Guillaume L, Gentzbittel L, Sarrafi A (2007) QTL mapping of partial resistance to Phoma basal stem and root necrosis in sunflower (*Helianthus annuus* L.). Plant Sci 172:815–823

Arnaud D, Dejardin A, Leple JC, Lesage-Descauses MC, Pilate G (2007) Genome-wide analysis of LIM gene family in *Populus trichocarpa*, *Arabidopsis thaliana*, and *Oryza sativa*. DNA Res 14:103–116

Bagge M, Lübberstedt T (2008) Functional markers in wheat: technical and economic aspects. Mol Breed 22:319–328

Barkley NA, Wang ML (2008) Application of TILLING and EcoTILLING as Reverse Genetic Approaches to Elucidate the Function of Genes in Plants and Animals. Curr Genomics 9:212–226

Barreiro LB, Henriques R, Mhlanga MM (2009) High-throughput SNP genotyping: combining tag SNPs and molecular beacons. Meth Mol Biol 578:255–276

Collard BCY, Cruz CMV, McNally KL, Virk PS, Mackill DJ (2008) Rice molecular breeding laboratories in the genomics era: current status and future considerations. Int J Plant Genomics 2008:524847

Comai L, Young K, Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner C, Odden AR, Henikoff S (2004) Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. Plant J 37:778–786

Cooper JL, Till BJ, Laport RG, Darlow MC, Kleffner JM, Jamai A, El-Mellouki T, Liu S, Ritchie R, Nielsen N, Bilyeu KD, Meksem K, Comai L, Henikoff S (2008) TILLING to detect induced mutations in soybean. BMC Plant Biol 8:9

Costabile M, Quach A, Ferrante A (2006) Molecular approaches in the diagnosis of primary immunodeficiency diseases. Hum Mutat 27:1163–1173

Dracatos PM, Cogan NO, Dobrowolski MP, Sawbridge TI, Spangenberg GC, Smith KF, Forster JW (2008) Discovery and genetic mapping of single nucleotide polymorphisms in candidate genes for pathogen defence response in perennial ryegrass (*Lolium perenne* L.). Theor Appl Genet 117:203–219

Ehrenreich I, Hanzawa Y, Chou L, Roe J, Kover P, Purugganan M (2009) Candidate gene association mapping of *Arabidopsis* flowering time. Genetics. doi:10.1534/genetics.109.105189

Elias R, Till BJ, Mba C, Al-Safadi B (2009) Optimizing TILLING and Ecotilling techniques for potato (*Solanum tuberosum* L). BMC Res Notes 2:141

Fusari CM, Lia VV, Hopp HE, Heinz RA, Paniego NB (2008) Identification of single nucleotide polymorphisms and analysis of linkage disequilibrium in sunflower elite inbred lines using the candidate gene approach. BMC Plant Biol 8:7

Galeano CH, Gomez M, Rodriguez LM, Blair MW (2009) CEL I nuclease digestion for SNP discovery and marker development in common bean (*Phaseolus vulgaris* L.). Crop Sci 49:1–14

Gilchrist EJ, Haughn GW, Ying CC, Otto SP, Zhuang J, Cheung D, Hamberger B, Aboutorabi F, Kalynyak T, Johnson L, Bohlmann J, Ellis BE, Douglas CJ, Cronk QCB (2006) Use of Ecotilling as an efficient SNP discovery tool to survey genetic variation in wild populations of *Populus trichocarpa*. Mol Ecol 15:1367–1378

Gilles PN, Wu DJ, Foster CB, Dillon PJ, Chanock SJ (1999) Single nucleotide polymorphic discrimination by an electronic dot blot assay on semiconductor microchips. Nat Biotechnol 17:365–370

Giordano M, Oefner PJ, Underhill PA, Cavalli Sforza LL, Tosi R, Richiardi PM (1999) Identification by denaturing high-performance liquid chromatography of numerous polymorphisms in a candidate region for multiple sclerosis susceptibility. Genomics 56:247–253

Haldane JB, Waddington CH (1931) Inbreeding and linkage. Genetics 16:357–374

Han W, Yip SP, Wang J, Yap MK (2004) Using denaturing HPLC for SNP discovery and genotyping, and establishing the linkage disequilibrium pattern for the all-trans-retinol dehydrogenase (RDH8) gene. J Hum Genet 49:16–23

Higuchi R, Fockler C, Dollinger G, Watson R (1993) Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. Biotechnology (NY) 11:1026–1030

Hung CC, Su YN, Lin CY, Chang YF, Chang CH, Cheng WF, Chen CA, Lee CN, Lin WL (2008) Comparison of the mismatch-specific endonuclease method and denaturing high-performance liquid chromatography for the identification of HBB gene mutations. BMC Biotechnol 8:62

Jones AC, Austin J, Hansen N, Hoogendoorn B, Oefner PJ, Cheadle JP, O'Donovan MC (1999) Optimal temperature selection for mutation detection by denaturing HPLC and

comparison to single-stranded conformation polymorphism and heteroduplex analysis. Clin Chem 45:1133–1140
Kolkman JM, Berry ST, Leon A, Slabaugh MB, Tang S, Gao W, Shintani DK, Burke JM, Knapp SJ (2007) Single nucleotide polymorphism and linkage disequilibrium in sunflower. Genetics 177:457–468
Kosambi DD (1944) The estimation of a map distance from recombination values. Ann Eugen 12:172–175
Kota R, Varshney RK, Prasad M, Zhang H, Stein N, Graner A (2008) EST-derived single nucleotide polymorphism markers for assembling genetic and physical maps of the barley genome. Funct Integr Genomics 8:223–233
Kota R, Wolf M, Michalek W, Graner A (2001) Application of denaturing high-performance liquid chromatography for mapping of single nucleotide polymorphisms in barley (*Hordeum vulgare* L.). Genome 44:523–528
Lai Z, Livingstone K, Zou Y, Church SA, Knapp SJ, Andrews J, Rieseberg LH (2005) Identification and mapping of SNPs from ESTs in sunflower. Theor Appl Genet 111:1532–1544
Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newberg LA (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181
Lexer C, Lai Z, Rieseberg LH (2003) Candidate gene polymorphisms associated with salt tolerance in wild sunflower hybrids: implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. New Phytol 161:225–233
Lijavetzky D, Cabezas JA, Ibanez A, Rodriguez V, Martinez-Zapater JM (2007) High throughput SNP discovery and genotyping in grapevine (*Vitis vinifera* L.) by combining a re-sequencing approach and SNPlex technology. BMC Genomics 8:424
Lin CH, Yeakley JM, McDaniel TK, Shen R (2009) Medium- to high-throughput SNP genotyping using VeraCode Microbeads. In: Bugert P (ed) DNA and RNA profiling in human blood: methods and protocols. Humana Press, pp 129–142
Liu A, Burke JM (2006) Patterns of nucleotide diversity in wild and cultivated sunflower. Genetics 173:321–330
Livak KJ, Flood SJ, Marmaro J, Giusti W, Deetz K (1995) Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. PCR Meth Appl 4:357–362
Makridakis NM, Reichardt JK (2001) Multiplex automated primer extension analysis: simultaneous genotyping of several polymorphisms. Biotechniques 31:1374–1380
Nieto C, Piron F, Dalmais M, Marco CF, Moriones E, Gómez-Guillamón ML, Truniger V, Gómez P, Garcia-Mas J, Aranda MA, Bendahmane A (2007) EcoTILLING for the identification of allelic variants of melon eIF4E, a factor that controls virus susceptibility. BMC Plant Biol 7:34
Oefner P, Underhill P (1998) DNA mutation detection using denaturing high-performance liquid chromatography (DHPLC). In: Current protocols in human genetics. Wiley, New York, pp 7.10.11–17.10.12. Supplement 19:7.10.1–7.10.12
Oleykowski CA, Bronson Mullins CR, Godwin AK, Yeung AT (1998) Mutation detection using a novel plant endonuclease. Nucleic Acids Res 26:4597–4602
Pajerowska-Mukhtar K, Stich B, Achenbach U, Ballvora A, Lubeck J, Strahwald J, Tacke E, Hofferbert HR, Ilarionova E, Bellin D, Walkemeier B, Basekow R, Kersten B, Gebhardt C (2009) Single nucleotide polymorphisms in the allene oxide synthase 2 gene are associated with field resistance to late blight in populations of tetraploid potato cultivars. Genetics 181:1115–1127
Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SP (1994) Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc Natl Acad Sci USA 91:5022–5026
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
Poormohammad Kiani S, Grieu P, Maury P, Hewezi T, Gentzbittel L, Sarrafi A (2007a) Genetic variability for physiological traits under drought conditions and differential expression of water stress-associated genes in sunflower (*Helianthus annuus* L.). Theor Appl Genet 114:193–207
Poormohammad Kiani S, Maury P, Sarrafi A, Grieu P (2008) QTL analysis of chlorophyll fluorescence parameters in sunflower (*Helianthus annuus* L.) under well-watered and water-stressed conditions. Plant Sci 175:565–573
Poormohammad Kiani S, Talia P, Grieu P, Maury P, Hewezi T, Gentzbittel L, Sarrafi A (2007b) Genetic analysis of plant water status and osmotic adjustment in recombinant inbred lines of sunflower under two water treatments. Plant Sci 178:773–787
Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100
Randall LP, Coldham NG, Woodward MJ (2005) Detection of mutations in *Salmonella enterica* gyrA, gyrB, parC and parE genes by denaturing high performance liquid chromatography (DHPLC) using standard HPLC instrumentation. J Antimicrob Chemother 56:619–623
Ren B, Zhou J-M, Komiyama M (2004) Straightforward detection of SNPs in doublestranded DNA by using exonuclease III/nuclease S1/PNA system. Nucleic Acids Res 34:e42
Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers [http://www.frodo.wi.mit.edu/]. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, NJ, pp 365–386
Sala CA, Bulos M, Echarte M, Whitt SR, Ascenzi R (2008) Molecular and biochemical characterization of an induced mutation conferring imidazolinone resistance in sunflower. Theor Appl Genet 118:105–112
Schiex T, Gaspin C (1997) CARTHAGENE: constructing and joining maximum likelihood genetic maps. Proc Int Conf Intell Syst Mol Biol 5:258–267
Shirasawa K, Maeda H, Monna L, Kishitani S, Nishio T (2007) The number of genes having diVerent alleles between rice cultivars estimated by SNP analysis. Theor Appl Genet 115:1067–1074
Southern EM (1996) High-density gridding: techniques and applications. Curr Opin Biotechnol 7:85–88
Spiegelman JI, Mindrinos MN, Oefner PJ (2000) High-accuracy DNA sequence variation screening by DHPLC. Biotechniques 29:1084–1090, 1092

Tang S, Kishore VK, Knapp SJ (2003) PCR-multiplexes for a genome-wide framework of simple sequence repeat marker loci in cultivated sunflower. Theor Appl Genet 107: 6–19

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

Till BJ, Colbert T, Tompa R, Enns LC, Codomo CA, Johnson JE, Reynolds SH, Henikoff JG, Greene EA, Steine MN, Comai L, Henikoff S (2003) High-throughput TILLING for functional genomics. In: Grotewold E (ed) Plant Functional Genomics: Methods and Protocols. Humana Press, Totowa, NJ, pp 205–220

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

Tsuchihashi Z, Dracopoli NC (2002) Progress in high throughput SNP genotyping methods. Pharmacogenomics J 2:103–110

Tyagi S, Kramer FR (1996) Molecular beacons: probes that fluoresce upon hybridization. Nat Biotechnol 14:303–308

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

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93:77–78

Wagner T, Stoppa-Lyonnet D, Fleischmann E, Muhr D, Pages S, Sandberg T, Caux V, Moeslinger R, Langbauer G, Borg A, Oefner P (1999) Denaturing high-performance liquid chromatography detects reliably BRCA1 and BRCA2 mutations. Genomics 62:369–376

Wang H-Y, Luo M, Tereshchenko IV, Frikker DM, Cui X, Li JY, Hu G, Chu Y, Azaro MA, Lin Y, Shen L, Yang Q, Kambouris ME, Gao R, Shih W, Li H (2005) A genotyping system capable of simultaneously analyzing > 1000 single nucleotide polymorphisms in a haploid genome. Genome Res 15:276–283

Wang J, Wang W, Liu Y, Duo L, Huang L, Jiang X (2009) The method of single-nucleotide variations detection using capillary electrophoresis and molecular beacons. Mol Biol Rep 36:1903–1908

Wang N, Wang Y, Tian F, King GJ, Zhang C, Long Y, Shi L, Meng J (2008) A functional genomics resource for *Brassica napus*: development of an EMS mutagenized population and discovery of FAE1 point mutations by TILLING. New Phytol 180:751–765

Weil CF, Monde R-A (2007) Getting the point—mutations in maize. Crop Sci 47:S60–S67

Xiao W, Oefner PJ (2001) Denaturing high-performance liquid chromatography: A review. Hum Mutat 17:439–474

Xin Z, Wang ML, Barkley NA, Burow G, Franks C, Pederson G, Burke J (2008) Applying genotyping (TILLING) and phenotyping analyses to elucidate gene function in a chemically induced sorghum mutant population. BMC Plant Biol 8:103

Yang W, Bai X, Kabelka E, Eaton C, Kamoun S, van der Knaap E, Francis D (2004) Discovery of single nucleotide polymorphisms in *Lycopersicon esculentum* by computer aided analysis of expressed sequence tags. Mol Breed 14:21–34

Yu B, Sawyer NA, Caramins M, Yuan ZG, Saunderson RB, Pamphlett R, Richmond DR, Jeremy RW, Trent RJ (2005) Denaturing high performance liquid chromatography: high throughput mutation screening in familial hypertrophic cardiomyopathy and SNP genotyping in motor neurone disease. J Clin Pathol 58:479–485