ORIGINAL PAPER

Evaluation of the reliability of maize reference assays for GMO quantification

**Nina Papazova · David Zhang · Kristina Gruden · Jana Vojvoda · Litao Yang ·
Meti Buh Gašparič · Andrej Blejec · Stephane Fouilloux · Marc De Loose ·
Isabel Taverniers**

Abstract A reliable PCR reference assay for relative genetically modified organism (GMO) quantification must be specific for the target taxon and amplify uniformly along the commercialised varieties within the considered taxon. Different reference assays for maize (*Zea mays* L.) are used in official methods for GMO quantification. In this study, we evaluated the reliability of eight existing maize reference assays, four of which are used in combination with an event-specific polymerase chain reaction (PCR) assay validated and published by the Community Reference Laboratory (CRL). We analysed the nucleotide sequence variation in the target genomic regions in a broad range of transgenic and conventional varieties and lines: MON 810 varieties cultivated in Spain and conventional varieties from various geographical origins and breeding history. In addition, the reliability of the assays was evaluated based on their PCR amplification performance. A single base pair substitution, corresponding to a single nucleotide polymorphism (SNP) reported in an earlier study, was observed in the forward primer of one of the studied alcohol dehydrogenase 1 (*Adh1*) (70) assays in a large number of varieties. The SNP presence is consistent with a poor PCR performance observed for this assay along the tested varieties. The obtained data show that the *Adh1* (70) assay used in the official CRL NK603 assay is unreliable. Based on our results from both the nucleotide stability study and the PCR performance test, we can conclude that the *Adh1* (136) reference assay (T25 and Bt11 assays) as well as the tested high mobility group protein gene assay, which also form parts of CRL methods for quantification, are highly reliable. Despite the observed uniformity in the nucleotide sequence of the invertase gene assay, the PCR performance test reveals that this target sequence might occur in more than one copy. Finally, although currently not forming a part of official quantification methods, *zein* and *SSIIb* assays are found to be highly reliable in terms of nucleotide stability and PCR performance and are proposed as good alternative targets for a reference assay for maize.

Nina Papazova and David Zhang contributed equally to this paper.

Electronic supplementary material The online version of this article (doi:10.1007/s00216-009-3386-4) contains supplementary material, which is available to authorized users.

N. Papazova · M. De Loose · I. Taverniers (✉)
Technology and Food Sciences,
Institute for Agricultural and Fisheries Research (ILVO),
Burg. Van Gansberghelaan 115,
9820 Merelbeke, Belgium
e-mail: Isabel.taverniers@ilvo.vlaanderen.be

D. Zhang · S. Fouilloux
GEVES Domaine du Magneraud, Laboratoire BioGEVES,
B.P. 52, 17700 Surgères, France

K. Gruden · J. Vojvoda · M. B. Gašparič · A. Blejec
Department of Biotechnology and Systems Biology,
National Institute of Biology (NIB),
Vecna pot 111,
1000 Ljubljana, Slovenia

L. Yang
GMO Detection Laboratory, SJTU-Bor Luh Food Safety Center,
School of life Science and Biotechnology,
Shanghai Jiao Tong University,
800 Dongchuan Road,
Shanghai 200240, China

Abbreviations

Adh1	Alcohol dehydrogenase 1 gene
BLAST	Basic Local Alignment Search Tool
Bp	Base pairs

CE	Capillary electrophoresis
CRL	Community Reference Laboratory
CRMs	Certified reference materials
CSCE	Conformation sensitive capillary electrophoresis
CSIC	Consejo Superior de Investigaciones Cientificas
CV	Coefficient of variation
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
EC	European Commission
EU	European Union
GM	Genetically modified
GMO	Genetically modified organism
Hmg	High mobility group protein gene
ILVO	Institute for Agricultural and Fisheries Research
Indel	Insertion/deletion
IRMM	Institute for Reference Materials and Measurements
Ivr	Invertase gene
JRC	Joint Research Centre
NIB	National Institute of Biology
PCR	Polymerase chain reaction
RA	Reference assay
SNP	Single nucleotide polymorphism
SSIIb	*Zea mays* starch synthetase type B gene

Introduction

The global area cultivated with biotech plant crops and the number of globally approved genetically modified (GM) crops and derived products are continuously growing [1]. In the European Union (EU) for example, controlling for the presence of genetically modified organism (GMOs) throughout the food and feed supply chains is necessary to be in compliance with a strict regulatory system [2–4]. A key aspect of this control system is the use by competent authorities of internationally validated, official polymerase chain reaction (PCR) methods for event-specific identification and relative quantification of each individual GM transformation event [3, 5–8].

In real-time PCR, a transgene event-specific assay and a target taxon or endogenous reference sequence assay are combined. The target taxon reference assay—further called reference assay (RA)—is used to quantify the total deoxyribonucleic acid (DNA) copy number and, further, to calculate the relative copy number proportion of the transgenic sequence. The functions of this RA in PCR-based GMO quantification determine the criteria which it has to fulfil: an RA should be taxon-specific, amplify a sequence which is present in a single copy number in the plant genome and amplify uniformly along the varieties within the target taxon.

Maize (*Zea mays* L.) has a high economical importance [9]. Maize cultured in the world can be divided into five types according to the grain characteristics: flint corn, dent corn, popcorn, sweet corn and waxy corn. Intensive breeding activities have led to many intermediate types between flint and dent corn. Originated in South America and cultured worldwide, the main gene pools involved in modern maize breeding activities are found in North America and Europe. Maize represents one of the most important biotech crops [1]. While only a few GM maize events have been approved in the EU for cultivation, more than 20 transgenic events have been approved so far for import or processing in food and/or feed products in the EU. Currently, 47 MON 810 maize varieties are approved for cultivation in Spain, France and Germany and registered in the Common catalogue of varieties [10].

For maize, various RAs have been developed and validated. The official Community Reference Laboratory (CRL) methods for quantification of different GM maize events employ different RAs, amplifying sequence parts of genes encoding for high mobility group protein gene (*Hmg*) and alcohol dehydrogenase 1 gene (*Adh1*) [8]. Other maize-specific methods detecting invertase (*Ivr*) and *zein* were developed and validated in-house or in the frame of collaborative projects [11, 12]. Another RA amplifying a region of the starch synthetase type B (*SSIIb*) gene is implemented in Japanese GMO standards [13–15]. Alternative maize RAs have been developed for reason to overcome possible modifications of the target genes such as copy number differences or mutations [11].

Maize-derived products which have to be tested for GMO presence may contain any of the maize varieties available on the market, which means that the maize RAs must perform uniformly along a broad range of plant varieties, transgenic or conventional. These varieties are produced by different breeding programmes utilising genetic resources from various origins. The maize genome is highly diverse. Single nucleotide polymorphisms (SNPs), duplications, deletions or insertions (indels) and transposon-like elements insertions are the sources of genetic variation determining the variability of different genomic loci between maize inbred lines [16–18]. SNPs are rather common in maize [16, 19, 20]. The average maize sequence diversity is much higher than the observed sequence variation for e.g. human and *Drosophila* genomes. Two randomly chosen maize loci vary on average in approximately one of 104 bases in contrast with the human genome where the random loci variation is on average one of 1,200–1,900 bases [20].

Variations in the nucleotide sequence in RAs' loci are likely to affect the GMO quantification. Single base pair substitutions or small indels could lead to primer or probe/template mismatching and further to a decrease in the

measured total DNA copy number [21–23]. An SNP leading to a decrease in the measured total DNA copy number was observed in the forward primer sequence of an internationally validated *Adh1* assay [21], demonstrating that although the method has been validated, the uniformity of the RA might not always be sufficiently assessed prior to implementation of the assay in routine GMO testing procedures.

This study aims at evaluating the reliability of eight existing maize RAs, four of which are part of official methods for GMO quantification. We analysed the nucleotide sequence variation in the target genomic regions in a broad range of transgenic and conventional varieties and lines: MON 810 varieties cultivated in Spain, conventional varieties and lines from various geographical origin and breeding history. Alternatively, the reliability of the assays was evaluated based on their PCR amplification performance in relation to the observed nucleotide variability.

Methods and materials

Maize samples and sample preparation

Seeds from 14 transgenic MON 810 varieties and five isogenic maize lines were kindly provided by Consejo Superior de Investigaciones Cientificas (CSIC; Spain). Seeds from 73 commercial varieties presenting several commercial and local breeding programmes were kindly provided by Agricultural Institute of Slovenia (Slovenia). Seeds from inbred maize line B73 were kindly provided by Bayer BioScience (Belgium). Seeds from inbred line HiII were kindly provided by European Commission-Joint Research Centre. Another set of 84 varieties was collected originating from Europe, South and North America, South Africa and China and covering the main types of use of maize (grain and forage corn, sweet corn, ornamental corn and popcorn). For additional data on the varieties included in the study, see the Table in the Supplementary material. Certified reference materials (CRMs) for maize GM events of MON 810, Bt176, Bt11, GA21, 1507, MON 863 and NK603 were purchased from the Institute for Reference Materials and Measurements (Geel, Belgium). These CRMs were certificated for GMO content (mass proportion, grammes per kilogramme). Transgene T25 maize material was supplied by Bayer CropScience (Gent, Belgium).

DNA extraction was performed using the DNeasy Plant Mini Kit (QIAGEN, Westburg) starting from 100 mg leaf material obtained from individual plants (Mon810 and conventional maize varieties). DNA was extracted from 250 mg of grain-derived powder as described in Zhang et al. [24] (84 varieties from the worldwide collection).

Maize target sequences

Eight maize RAs targeting five different maize-specific genes were considered in this study (Table 1, A). For *Adh1*, three publicly available target sequences are considered: one used as RA in the CRL methods for quantification of NK603, MON863 and GA21 (2005 assay, GA21^{2005}) maize, and two others used in the CRL methods for T25, respectively, Bt11, GA21 (2007 assay, GA21^{2007}), MIR604 and 3272 maize. The latter two assays are more or less one and the same assay, expect that the forward primer for Adh1 in the T25 assay has a single base more in length than the forward primer in the RA for Bt11. Further, in this paper, these two assays will be considered as one and the same. For the purpose of this paper, two different RAs for the same endogenous gene will be distinguished by adding the amplicon length between brackets: the two *Adh1* assays will be named *Adh1* (70) and *Adh1* (136), respectively, while the two zein assays will be named zein (104) and zein (110), respectively (Table 1, A).

PCR analyses

Unless otherwise stated, all apparatuses, software and specific reagents are from Applied Biosystems. PCRs for conformation sensitive capillary electrophoresis (CSCE) as well as for sequence analysis were performed in 25 μl volumes containing approximately 100 ng DNA, 300 nM from both forward and reverse primers, 200 μM deoxynucleotide triphosphates and 1 U AmpliTaq DNA polymerase. Thermal cycling reactions were performed under the following conditions: initial denaturation at 94 °C for 3 min, followed by 35 cycles of denaturation at 94 °C for 30 s, primer annealing at 57 or 59 °C (depending on the primer pair) for 40 s, and elongation at 72 °C for 40 s to 1 min depending on the amplicon size. The final extension reaction was performed at 72 °C for 5 min, after which, the reaction was held at 4 °C. The PCR products were used either for CSCE analysis or for sequencing. The products used for CSCE were amplified using 3′FAM- or 3′HEX-labelled primers, while for sequencing, the products were amplified with non-labelled primers (see Table 1, B). Before sequencing, the PCR products were purified by means of MSB® Spin PCRapace (Invitek, Belgium). Where appropriate, primers were designed by means of Primer3 v. 0.3.0 [25]. All primer sequences are shown in Table 1. To control the success of the amplification reactions, PCR products were analysed using agarose gel electrophoresis.

CSCE analysis

Pools of DNA were made by combining DNAs from four to eight lines in equal concentrations. The pooled DNAs were

Table 1 Published maize reference assays (RAs; *A*) and used oligonucleotides in polymerase chain reaction (PCR) analyses for the different purposes described in this study: conformation sensitive capillary electrophoresis (CSCE) analysis (*B*), sequencing analysis (*B*) and real-time PCR analysis (*A*)

Zea mays gene	Oligo name	Use	5′–3′ sequence	Amplicon (bp)	Reference
A					
Adh1[a]	Adh primer F	Forward	CCA GCC TCA TGG CCA AAG	70	*NK603, MON863, GA21*[2005]
	Adh primer R	Reverse	CCT TCT TGG CGG CTT ATC TG		
	Adh probe	Probe	CTT AGG GGC AGA CTC CCG TGT TCC CT		
Adh1	Adh-F3	Forward	CGT CGT TTC CCA TCT CTT CCT CC	136	*Bt11, GA21*[2007]*, MIR604,*
	Adh-R4	Reverse	CCA CTC CGA GAC CCT CAG TC		*3272,* [11]
	Adh1-MDO	Probe	AAT CAG GGC TCA TTT TCT CGC TCC TCA		
Adh1[a]	KVM182	Forward	CGT CGT TTC CCA TCT CTT CCT CCT	136	*T25*
	KVM183	Reverse	CCA CTC CGA GAC CCT CAG TC		
	TM014	Probe	AAT CAG GGC TCA TTT TCT CGC TCC TCA		
Ivr1[b]	Ivr7	Forward	CGC TCT GTA CAA GCG TGC	104	[11]
	Ivr8	Reverse	GCA AAG TGT TGT GCT TGG ACC		
	Ivrp	Probe	CAC GTG AGA ATT CCC GTC TAC TCG AGC CT		
Hmg[b]	Mhmg for	Forward	TTG GAC TAG AAA TCT CGT GCT GA	79	*TC1507, DAS59122, MON88017/*
	Mhmg rev	Reverse	GCT ACA TAG GGA GCC TTG TCC T		*89034/810, LY038,* [11]
	Mhmg probe	Probe	CAA TCC ACA CAA ACG CAC GCG TA		
Zein[a]	taqze1	Forward	GCC ATT GGG TAC CAT GAA CC	104	[11]
	taqze3	Reverse	AGG CCA ACA GTT GCT GCA G		
	taqze2	Probe	AGC TTG ATG GCG TGT CCG TCC CT		
Zein[a]	ENDOZEIN-1	Forward	GCA TTG TTC GCT CTC CTA GC	110	[12]
	ENDOZEIN-2	Reverse	TAC TGC ATC CAT GGG TTC AT		
	TQ-ZEIN	Probe	TTC CAG GGC ACT TGC CAC CA		
SSIIb[a]	SSIIb 1-5′	Forward	CTC CCA ATC CTT TGA CAT CTG C TCG ATT	114	[13–15]
	SSIIb 1-3	Reverse	TCT CTC TTG GTG ACA GG		
	SSIIb-Taq	Probe	AGCAAAGTCAGAGCGCTGCAATGCA		
B					
CSCE					
Adh1	Adh-csce-f1	Forward	ATC GAG GAG GTG GAG GTA GC	395	This paper
	Adh-csce-r1	Reverse	TCT TCC ACA TGC ATC TAG TAA AAC A		
Zein	Zein-csce-f1	Forward	CTC TAG GAA GCA AGG ACA CCA	354	This paper
	Zein-csce-r1	Reverse	GCA TCA TCA TTT GCG ACA TC-FAM		
SSIIb	SSIIb-csce-f1	Forward	GTG ACA CAC GCG CTT GTA TAT T-HEX	350	This paper
	SSIIb-csce-r1	Reverse	AGG TTC ATT ATC CTC CCT GTC AT		
Sequencing					
Adh1	Adh-seq-f1	Forward	CGT CTA GAA ATC CTA ATG ATT CAG C	576	This paper
	Adh-seq-r1	Reverse	GAT CAA GAG ACA GAG GAG AAA CAA G		
Adh1	Adh-seq-f2	Forward	TTG AAT TTA ACT AAC TCG TTG AGT GG	657	This paper
	Adh-seq-r2	Reverse	CGT ATG TCC ACA CAG ATA ATA TCC A		
Zein	Zein-seq-f1	Forward	CTC TAG GAA GCA AGG ACA CCA	354	This paper
	Zein-seq-r1	Reverse	GCA TCA TCA TTT GCG ACA TC		
SSIIb	SSIIb-seq-f1	Forward	GTG ACA CAC GCG CTT GTA TAT T	634	This paper
	SSIIb-seq-r1	Reverse	AGG TTC ATT ATC CTC CCT GTC AT		

All TaqMan probes mentioned are 5′ FAM-labelled and 3′ TAMRA-labelled. GM events mentioned under "Reference" in italics refer to RAs as used in the official CRL-GMFF event-specific methods available from [8]. In case if two CRL methods are available for one event, the year of publication is added (see GA21 assays)

[a] Those RAs were used for real-time PCR analyses

[b] RAs used both for the CSCE as well as for the real-time PCR analyses

used as a template for PCR amplification with the primers targeting the regions of interest (Table 1, B, for *Ivr1* and *Hmg*, see Table 1, A). DNA of inbred line B73 was added to each pool as a reference. After the PCR amplification, heteroduplexes were formed by denaturation followed by re-hybridization under the following conditions: 96 °C for 10 min; 72 cycles (96 °C, 60 °C (−0.5 °C/cycle) for 20 s), 60 °C for 30 min, then cooled down to room temperature. The homo- and heteroduplexes were electrophoretically separated at 18 °C for 2,500 s, at voltage 10 kV using an ABI3130 Genetic Analyzer. The instrument has a capacity of 96 samples automatically injected at 10 kV for 10 s to a 60-cm long, 50-μm diameter capillary filled with the sieving matrix Conformational Analysis Polymer (CAP®). This polymer is composed of 5% native polymer diluted with TTE glycerol tolerant buffer, 4 M urea and 4% sucrose. The relative position of the fluorescent peaks was determined by comparison with an internal 500 ROX Standard. The level of fluorescence was monitored by diluting the samples 120 up to 1,000 times. The fluorescent peak patterns were analysed by using GeneMapper 3.7. All samples were run in duplicates.

Sequencing reactions

Sequencing reactions were performed using the BigDyeTM Deoxy Terminator Sequencing Kit v.1.1 according to the manufacturer's instructions on an ABI3130-Avant Genetic Analyzer system. For the lane tracking and first pass base calling, Sequencing Analysis software v.5.2 was used. For the SNP characterisation, the sequences were assembled and aligned against the respective reference sequence by means of SeqScape 2.5.

Real-time PCR analyses

DNA concentrations were estimated using the PicoGreen® ds DNA quantification kit (Molecular Probes, The Netherlands), and the FLx800 microplate fluorescence reader (Bio-Tek® Instruments, Inc., Winooski, VT, USA). DNA quantities were related to a calibration curve made of four points (0.2, 2, 20 and 200 ng Lambda DNA). Real-time PCRs amplifying *Adh1* (70), *Adh1* (136), *Ivr*, *Hmg*, *zein* (110) and *SSIIb* target sequences were performed as described in original publications (see Table 1, A). All DNA samples from the worldwide variety collection (84 samples) were diluted to the same two concentrations (1 and 0.125 ng/μl). These two sets of dilutions were used as templates in the different real-time PCR systems. Measured Ct values for five endogenous RAs were obtained for both sets of dilutions (total 168 samples) and used for further analysis. Calibration curves for maize DNA copy number quantification were made of

four points: 60,000>15,000>3,750>937.5 haploid T25 maize genomes (100%). For quantifying the GM DNA, the calibration curves were made of five points: 1,600>800>200>50>2.5 haploid genomes. The calculations of haploid genome copy number were based on extrapolation from the 1C mass of maize: 1 ng of maize DNA equals to approximately 385 maize haploid genomes [26]. For evaluation of the SNP identified in the *zein* gene sequence, four samples were selected giving different results of sequencing analysis and analysed in five dilutions for all RAs. All samples were analysed in at least two parallels of each DNA dilution.

Statistical analysis of real-time PCR data

For each reference system, a standard curve was constructed by plotting Ct values against $\log_{10}$ of DNA copy numbers of known standards and fitted by linear least square regression. The amplification efficiency in the individual sample was examined by calculating the slopes using the Ct values of different dilutions (1 and 0.125 ng/μl) and the log copy number values. Samples that show bad amplification efficiency in all RAs were completely excluded from further evaluation. Ct values of both sets of dilutions were converted to copy numbers using standard curves. The GeNorm software was used to estimate the stability of the RAs. Copy numbers were used as the input data according to the request of GeNorm. The application automatically calculates a "gene-stability measure" M for all investigated genes in a given set of samples. The parameter M is defined as the average pair-wise variation in measured copy numbers of a particular gene with all other genes of interest [27]. The RA is considered more or less reliable based on the M value; RAs with the lowest M value are more stable than those with higher M value. A genewise normalisation of copy numbers was performed in datasets of both dilutions. The highest copy number value was used as the normalisation factor (highest copy was set to 1). Normalised data were then visualised in two different ways using "R", a free software environment for statistical computing and graphics. Scatter-plots were constructed for each pair of RAs to estimate stability of target copy number estimates using an independent approach; copy numbers obtained in analyses of one gene were plotted against copy numbers obtained in analyses of another gene. Boxplots were constructed to allocate the variability in performance of RAs to individual samples. An outlier in copy number estimation is defined as a true positive, when the overall sample variability in copy number obtained by different RAs was larger than average and when the efficiency of amplification of the RA was less than 15% deviating from the efficiency of amplification of reference material.

Results

Sequence variation analysis through CSCE screening

Screening of the *Ivr*, *Hmg* and *SSIIb* target taxon genes by means of CSCE revealed presence of a single major peak in all of the examined pools. The peaks occurred at positions corresponding to 120, 190 and 380 bp, respectively, for the three targets. The position of the peaks was consistent in all the pools and in all dilutions of the PCR fragments. The presence of a single or of multiple peaks was identified based on the analysis of samples with optimal dilution e.g. displaying fluorescent peaks comparable in height with this of the size marker fragments (Fig. 1). The presence of a single clear and sharp peak corresponds to the homoduplex, which means absence of nucleotide variation in the sequences of the analysed *Ivr*, *Hmg* and *SSIIb* amplicons present in the pools. Amongst the individual genotypes, no nucleotide variation was observed for the conventional varieties and lines as well as for MON 810 varieties.

Multiple peaks were observed in the *Adh1* and *zein* patterns of some of the pools, while in other pools, a single major peak corresponding to the homoduplex was ob-
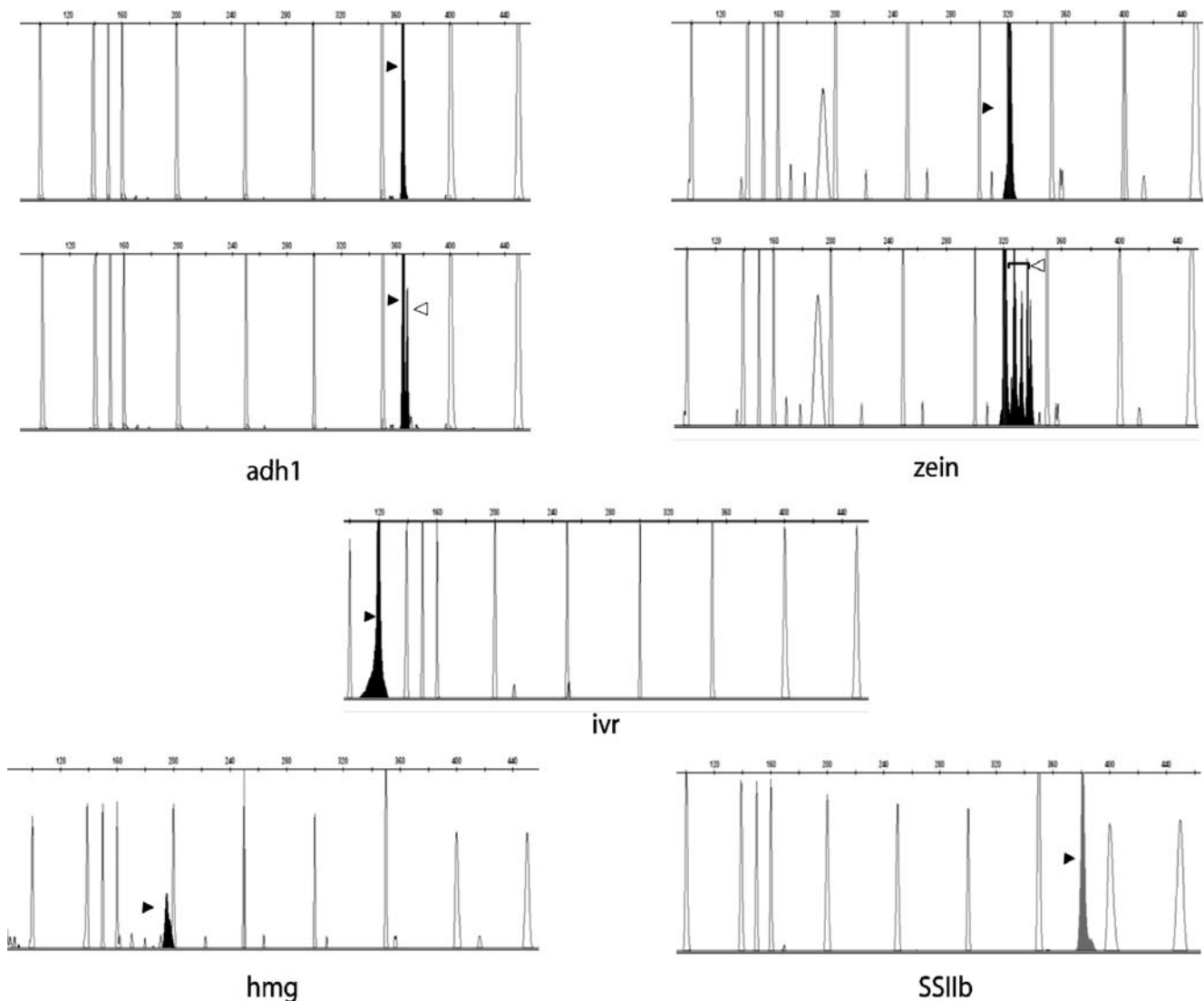

adh1

zein

ivr

hmg

SSIIb

Fig. 1 Results of conformation sensitive capillary electrophoresis screening for the presence of single nucleotide polymorphisms within a pool of relevant polymerase chain reaction fragments, shown as peaks (units of fluorescence, *y*-axis) in function of the fragment length (*x*-axis). The fluorescent peaks of the respective fragments are *black* (FAM-labelled fragments) or *grey* (HEX-labelled fragments). The major peaks corresponding to the heteroduplexes are indicated with a black arrow, the minor peaks indicating the presence of heteroduplexes (where present) are indicated with a *transparent arrow*. The transparent peaks correspond to the size marker. The length of the size marker is per step of 50 base pairs (bp), starting with 100 bp (first length marker peak, X-axis on the left) and ending with 450 bp (last length marker peak, X-axis on the right)

served. The position of the peaks was consistent in all analysed pools. The major peaks corresponding to the homoduplex occurred at approximately 365 and 320 bp for the *Adh1* and *zein* target sequences, respectively (Fig. 1, upper patterns of *Adh1* and *zein* panels). A second minor peak occurring in the proximity of the major one was identified in the *Adh1* pattern of some of the pools (Fig. 1, lower pattern of *Adh1* panel). In the *zein* pattern, four minor peaks with lower mobility than the major one were observed (Fig. 1, lower pattern of *zein* panel). The presence of more than one fluorescent peak in the CSCE analysis indicates the presence of nucleotide variation between the individuals present in the pool. In case of heteroduplex formation, four peaks should normally be observed. However, sometimes the heteroduplexes migrate together with the homoduplexes and cannot be sufficiently distinguished. Presence of more than one peak is usually sufficient to conclude that nucleotide variation exists [28]. The individuals from the pools displaying the presence of a heteroduplex were further analysed by sequencing to confirm and characterise the observed nucleotide variation. In addition, one individual sample from each pool displaying a single peak was analysed to determine the type of the homoduplex sequence. This has been done for both *Adh1* and *zein* fragments.

Characterization of observed SNP alleles

SNP alleles in Adh1 (70) in conventional and MON 810 varieties

Nucleotide sequences of the individuals from the pools displaying presence of multiple peaks were analysed. Sequence alignments were done against the reference sequence

corresponding to the F allele of *Adh1* (X04049) [29]. The alignments revealed that the nucleotide sequence for 39 of the analysed genotypes corresponds to the F allele of *Adh1*. In four of the lines, a heterozygous SNP C/T at position 2039 was observed. In one of the analysed individuals, a homozygous SNP (C>T) was observed (Table 2).

Sequencing analysis of the individuals from the pools displaying a single peak revealed that the homoduplex corresponds to the *Adh1* F allele. The observed homozygous SNP (C>T) in one out of 44 sequenced genotypes matches the S allele of *Adh1* [29].

Sequencing alignment of the obtained sequences and primers used in real-time PCR showed that the single base pair substitution is located in the forward primer attachment site at position +9 (Fig. 2). The observed SNP corresponds to the one described by Broothaerts et al. [21].

The nucleotide variation of the considered *Adh1* amplicon sequence was studied in more detail in the MON 810 varieties. The sequence analysis revealed that five of the transgenic lines and four of the corresponding isogenic counterparts contain the sequence corresponding to the F allele. In five of the transgenic lines and one of the isogenic lines, the heterozygous SNP allele was observed. None of the varieties contained the homozygous S allele (Table 2). Sequencing analysis was performed on the regions flanking the *Adh1* target sequence. The analysis further confirmed that these lines carry the F allele of *Adh1*. No other nucleotide variation was observed amongst the lines along the *Adh1* sequence. The analysed region also covered the region of the transcription initiation, which was used for the design of two other real-time PCR *Adh1* assays [11]. The region was uniform among the varieties, and no polymorphisms were found between the lines and the reference sequence (Fig. 2).

Table 2 Summary of the observed and characterised single nucleotide polymorphisms (SNPs) related to different reference assays in the transgenic and conventional maize varieties and lines

Reference assay	Total number of lines analysed	Number of varieties tested/number and type of SNP alleles observed		
		Mon 810	MON 810 isogenic lines	Conventional varieties and lines
Adh1 (70)	102	5/C[a]; 5/C/T[b]	4/C[a]; 1/C/T[b]	39/C[a]; 4/C/T[a]; 1/T[b]
Adh1 (136)	18	13/0	5/0	ND
Hmg	174	14/0	5/0	155/0
Ivr	170	14/0	5/0	151/0
zSSIIb	174	14/0	5/0	155/0
Zein (104)	161	14/0	5/0	142/0
Zein (110)	171	2/G[a]; 1/C[a]; 9/G/C[b]	1/G[a]; 2/C[a]; 2 G/C[b]	18/G[a]; 84/C[a], 16 G/C[b]

ND not determined

[a] Homozygous SNP allele
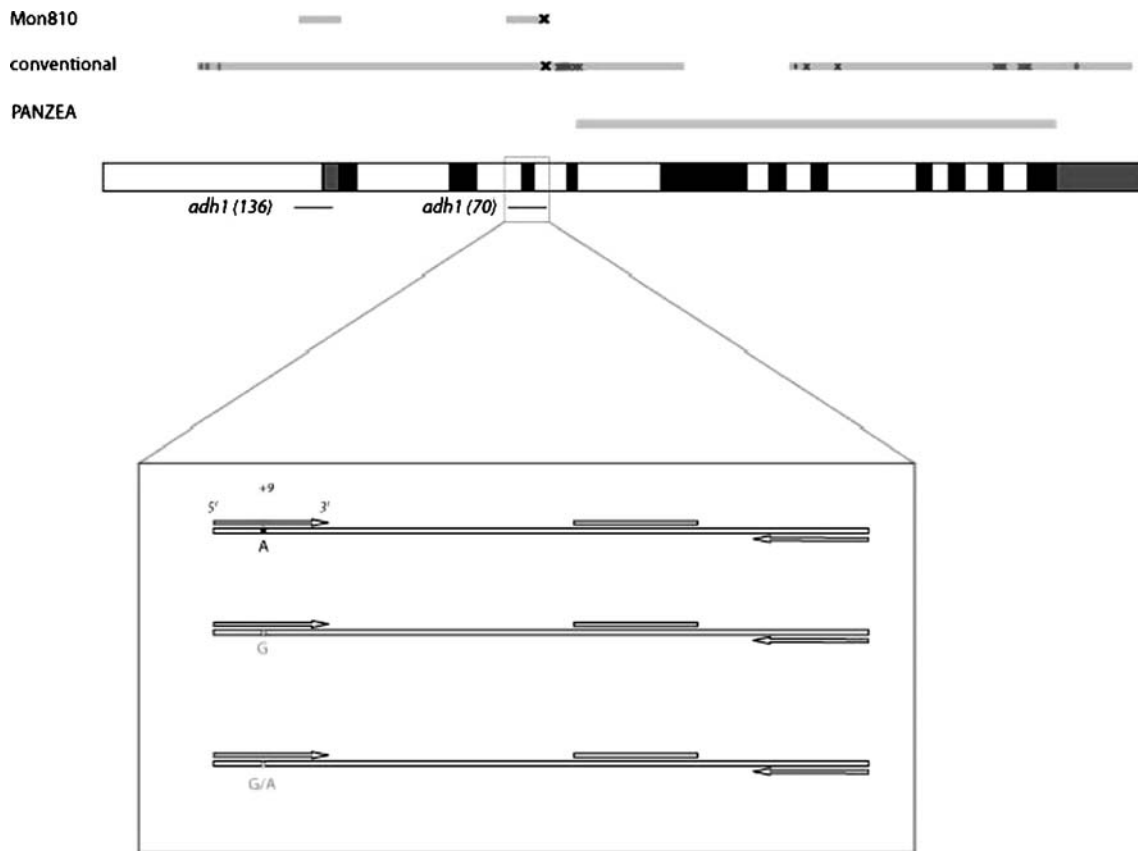
[b] Heterozygous SNP allele

Fig. 2 Schematic representation of single nucleotide polymorphisms along the *Adh1* sequence and location of the single nucleotide polymorphism (SNP) along the sequence of the Adh1 (70) assay. The structure of the F allele of Adh1 is presented according to Sachs et al. [29]. PANZEA refers to nine inbred lines where 38 candidate SNPs were observed in the *Adh1* sequence (available from http://www.panzea.org). The base pair substitution observed in this study is presented with an *x* along the gene sequence. The *black cross*

indicates the SNP occurring in the Adh1 (70) assay and observed in the lines tested in this study. The location of the SNP along the real-time polymerase chain reaction Adh1 (70) assay is shown in details in the figure below. The forward primer sequence is reverse complement to the S allele of *Adh1* (X04050) and the base pair in the forward attachment site at 5′+9 is A (*upper figure*). A single base pair substitution A>G or A>G/A occurs in this position of the primer (*lower figure*). The scale is proportional

SNP alleles in zein (110) sequence in conventional and transgenic varieties

The nucleotide sequence of the *zein* fragment of the individual samples from the pools displaying multiple peaks was further characterised by sequencing. Sequencing alignments were performed against the sequence corresponding to the AF371266 allele. The alignments revealed that 84 of the conventional varieties correspond to this allele. In 18 of

the analysed genotypes, a homozygous base pair substitution (C>G) was observed. In 16 of the genotypes, the heterozygous SNP allele was detected. The three genotypes were observed in the MON 810 and the isogenic lines as well (Table 2). A Basic Local Alignment Search Tool located the SNP at position 125 of AF371266. This corresponds to the regions between the forward primer and the probe attachment site of the *zein* (110) assay. The sequence of the *zein (104)* assay remained uniform amongst the examined varieties (Fig. 3).

Fig. 3 Schematic representation of the observed single nucleotide polymorphisms in the *zein* target sequence along the real-time polymerase chain reaction (PCR) assays. The *grey arrows and lines* represent the primers and probe attachment sites of the zein (110) real-

time PCR assay. The *white arrows and lines* represent the primers and probe attachment sites of the zein (104) assay. The observed single nucleotide polymorphism is presented with a *black box* on the DNA template

Table 3 Real-time polymerase chain reaction amplification of the six maize reference systems on 84 maize cultivars representing different geographical origins

Reference gene system	Normalised copy number mean ± SD	Difference between extreme values	Mean of PCR efficiency (%)
Adh1 (70)	0.34±0.21	0.92	96.54
Adh1 (136)	0.51±0.15	0.81	96.7
Hmg	0.57±0.14	0.73	89.46
SSIIb	0.41±0.15	0.86	80.77
Zein (110)	0.57±0.15	0.71	84.84
Ivr1	0.72±0.14	0.62	94.24

Normalised copy numbers were used for calculations

Real-time PCR performance testing of reference assays on worldwide collection of varieties

To assess the overall performance of different RAs (Table 1, A), the collection of 84 lines was used. First, comparison of variability in estimated copy numbers and variability in efficiency of amplification was evaluated (Table 3). In the raw data, the difference between extreme Ct values was less than 1 Ct in case of *Hmg* and *Adh1* (136) while higher than 2 Cts in *Ivr* and *Adh1* (70) RAs. Further, the RA stability in real-time PCR amplification was evaluated using GeNorm statistics. *Adh1* (70) gives the highest *M* value, showing low stability of this RA. The variability of amplification in *Adh1* (136), *zein* (110), *Ivr* and *Hmg* assays was low in the analysed varieties (Fig. 4).

After the genewise normalisation of copy number data, scatter plots were constructed for each pair of assays to show the correlation between copy numbers obtained with two different RAs. Best correlation of results was obtained when using *zein* (110), *Hmg* and *SSIIb* RAs. Extremely

dispersed scatter plots and correlation coefficients of less than 0.5 were calculated for *Adh1* (70) RA copy number estimations versus all other RAs included in the analysis (Fig. 5).

Boxplots of normalised copy numbers in samples were constructed to allocate the variability in performance of the RAs to individual samples. Outliers found using this approach can be candidates for genotypes with sequence modification in the primer/probe target region or changes in target copy number in the genome. Following our control parameters, outliers were identified in *Adh1* (70) and *Ivr* determined copy numbers only. *Adh1* (70) was a negative outlier, meaning that numbers of copies are always underestimated, and *Ivr* was a positive outlier with overestimated copy numbers. In two samples (varieties 43 and 62), outliers for *Ivr* determined copy numbers were identified indicating potential duplication of the *Ivr* gene in those genotypes. In five samples (varieties 2, 36, 49, 61, 62 and 72), *Adh1* (70) was identified as an outlier in target copy number estimation. No outliers were detected using our criteria in copy number estimation performed by other RAs.

In order to know whether the variability in performance of the RAs depends on the geographical origins (gene pools) of varieties, 73 maize cultivars used for grain production were grouped into five geographical origins corresponding to the main regions of maize culture in the world: South Africa, Asia, South America, North America and Europe. Table 4 shows that the difference of the mean normalised copy numbers between geographical groups is statistically non-significant ($p<0.05$) in all analysed RAs if considering only samples with acceptable efficiency of amplification. These results prove our expectations that there is less variation related to the geographical origins (gene pools) than to RA groups. Similar analysis was performed with varieties grouped according to their use: grain and forage corn, sweet corn, ornamental and popcorn (Table 4). Also, in this setup, the differences in copy number estimations between types of use are insignificant for all RA considered.
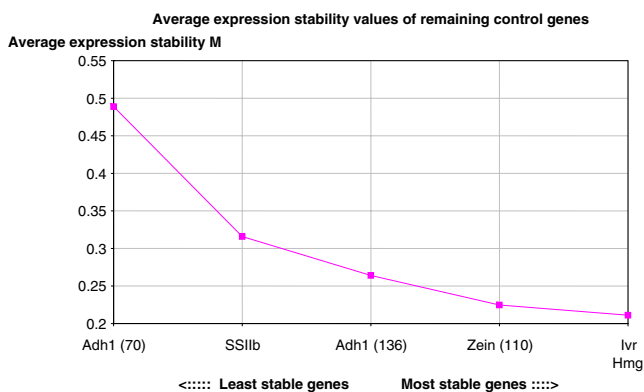

Fig. 4 Stability of reference assays amplification as evaluated in GeNorm. Average stability measure *M* (*y*-axis) was calculated for all tested reference assays, plotted on the *x*-axis. Results from the analyses of the 1 ng/µl DNA dilution set

Fig. 5 Comparison of performance of reference assays in worldwide collection of varieties. Each scatter plot was constructed by plotting the copy numbers determined with one reference assay (x-axis) against the copy numbers determined with another reference assay (y-axis) after the genewise normalisation of copy number data. The red line in each scatter plot denotes a perfect correlation (r=1) of copy numbers between the two assays. The individual correlation coefficients between the assays are also shown in the squares diagonally opposite to individual plots
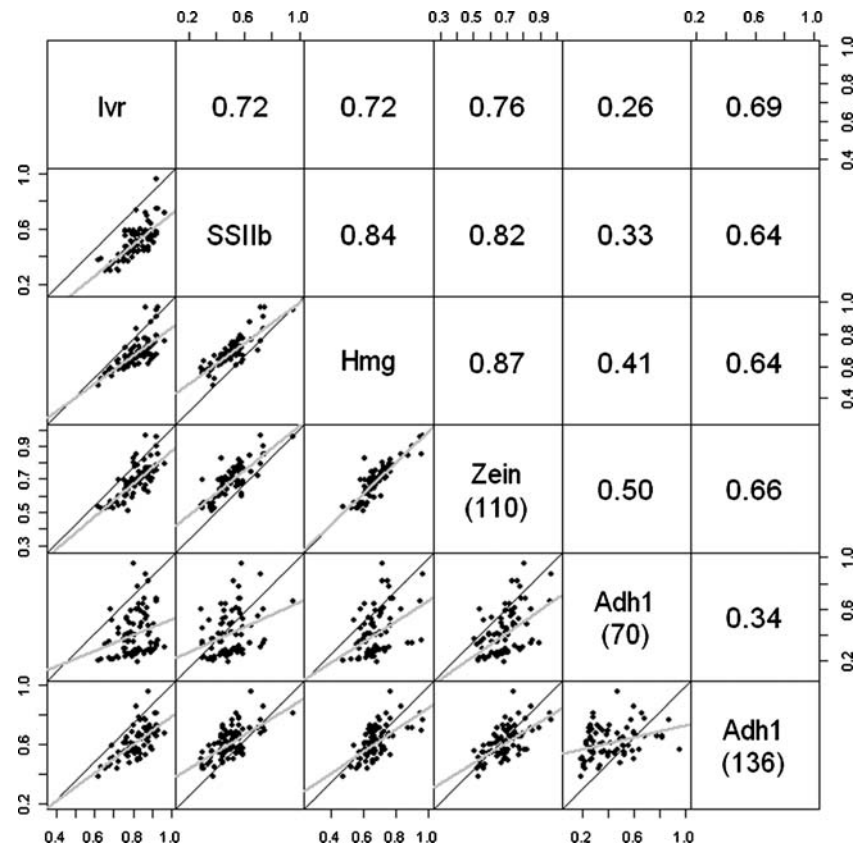
Discussion

We aimed at studying the reliability of a number of existing maize reference PCR assays which are widely used in GMO testing labs. A reliable RA must be specific for the taxon of interest, uniform along the varieties within the taxon and with low and stable copy number. The specificity of maize RAs was already assessed *in silico* and experimentally prior to method validation [8]. The uniformity was our main concern for several reasons. First, the maize genome is highly diverse, and SNPs are very common. SNPs occur between the maize inbred lines, which are

Table 4 Mean copy numbers and the extreme copy number difference observed on the five maize reference gene systems according to geographical origin and grain utility

Geographical origin or type of maize	Number of cultivars	Adh1 (136) Mean copy number ±SD	Adh1 (70) Mean copy number ± SD	*Hmg* Mean copy number ± SD	SSIIb Mean copy number ± SD	Zein1 Mean copy number ± SD	*Ivr* Mean copy number ± SD
According to the geographical origin among the grain and forage corn							
South Africa	7	0.65±0.10	0.33±0.17	0.58±0.17	0.42±0.16	0.61±0.16	0.73±0.15
China	7	0.66±0.23	0.46±0.27	0.68±0.17	0.54±0.14	0.69±0.16	0.80±0.07
South America	10	0.62±0.09	0.58±0.22	0.67±0.12	0.50±0.18	0.71±0.11	0.83±0.06
North America	16	0.45±0.14	0.28±0.17	0.57±0.19	0.39±0.18	0.53±0.19	0.71±0.19
Europe	30	0.47±0.10	0.28±0.19	0.50±0.08	0.36±0.13	0.51±0.13	0.69±0.13
According to the type of use							
Grain and forage corn	70	0.52±0.15	0.34±0.22	0.57±0.15	0.41±0.16	0.57±0.16	0.73±0.15
Sweet corn	5	0.49±0.17	0.35±0.20	0.52±0.05	0.36±0.07	0.53±0.04	0.70±0.06
Ornamental	5	0.42±0.04	0.28±0.17	0.58±0.04	0.43±0.09	0.57±0.10	0.69±0.05
Popcorn	2	0.47±0.08	0.28±0.11	0.57±0.03	0.44±0.03	0.57±0.001	0.65±0.04

Normalised copy numbers were used for calculations

frequently used in breeding programmes. Second, RAs used for the calculation of the relative GMO content in a sample have to amplify uniformly along the large number of commercially available varieties, which might be present in test samples of either seeds, grains, food or feed. In the EU only, more than 1,000 maize varieties are registered [10], and additional maize varieties might be imported from third countries [30]. The nucleotide stability of the RAs genome targets along the abundance of commercial varieties might not be sufficiently assessed prior to method validation. Variation in the assays' performance along the varieties might be expected when the assay has to be applied at the analytical lab to analyse unknown samples. This means that some RAs might perform better than other. The commercially cultivated varieties are produced in different years, by different breeders. Given the natural base pair substitution rate of coding maize genome sequences (10^{-8} bp) [31], it is unlikely that for the production cycle of one variety (approximately 10 years), a significant number of nucleotide substitutions would accumulate in the maize genome potentially leading to non-uniformity of the nucleotide sequence especially in genes used for design of a RA. However, pre-existing nucleotide variation due to the presence of different SNP alleles among the breeding lines could give rise to variation in the nucleotide sequence among the varieties.

The uniformity of the maize RAs was studied in two manners: (1) straightforwardly, by analysing the nucleotide sequence of targeted regions, and (2) indirectly, by statistical analysis of the real-time PCR derived parameters. The first approach gives a direct view on the variation of the nucleotide sequences of the target regions and allows to locate an SNP, if any, along the primer and probe attachment site. The second approach evaluates the uniformity of the PCR performance along the varieties and additionally indicates if the RA amplification is positively or negatively affected. It is used to indicate either nucleotide variability or unstable copy number of the targeted gene.

Reference assays for official GMO control purposes are developed and tested on a limited number of varieties which, in some cases, can result in RAs with lower uniformity [21]. Here, we studied two groups of maize plant materials: conventional commercial varieties and lines bred in different parts of the world on one hand and commercial transgenic lines (MON 810) cultivated currently in Spain on the other hand. The collections of varieties included a wide number of commercial varieties and lines available at the market, representing the variability within the maize botanical taxon.

The nucleotide sequence analysis shows that four of the tested RAs (*Adh1* (136), *Ivr*, *Hmg* and *SSIIb*) have genome targets uniform in their nucleotide sequence in the tested varieties (see Table 2). However, from these four assays, the *Hmg* assay seems to be the most stable, displaying consistent PCR performance and nucleotide sequence uniformity. This complies also with the data obtained by Broothaerts et al. [21] where *Hmg* showed low variation in the quantified copy number (coefficient of variation) along four maize conventional varieties and seven different transgenic events. The *Hmg* real-time PCR assay which forms a part of 14 official CRL methods for GMO quantification can be qualified as highly reliable.

The *Ivr* RA did not indicate non-uniformity of the nucleotide sequence; however, this assay displayed non-consistent PCR results, indicating that the target sequence might occur in more than one copy. Similar suggestions were done by Moor [32]. *Ivr* displayed overestimated and highly variable copy numbers along a number of varieties. However, the author could not conclude if the sequence exists in multiple copies, and any other literature references were not available.

The *SSIIb* RA displayed a uniform sequence and relatively good real-time PCR performance. The scattered plots of *SSIIb* compared to the rest of the tested RAs were not dispersed. However, *SSIIb* performed slightly worse in GeNorm analysis. Broothaerts et al. [21] demonstrated low variability in the amplification of the *SSIIb* target sequence comparable to the one of *Hmg*. Taken altogether, we could consider this method as reliable.

Nucleotide variation was found in the tested target sequences of the *Adh1* (70) and *zein* (110) assays. The nucleotide variation observed in the *zein* (110) sequence was not located at primer or probe attachment sites, suggesting that no influence on the real-time PCR amplification might be expected. This was consistent with the real-time PCR results displaying a stable performance comparable to this of *Hmg*. *Zein* (110) assay partially overlaps with *zein* (104) assay previously designed by Hernandez et al. [11]. These authors already suggested that *zein* might occur in more than one copy in some lines. We did not observe any indication that this sequence might have higher copy number. The scattered plots were not dispersed, and the observed variation was very low. *Zein* RAs are not a part of CRL validated methods. However, both assays are reliable in terms of sequence uniformity.

The single base pair substitution observed in the primer attachment site of *Adh1* (70) corresponds to the one reported by Broothaerts et al. [21]. The authors observed that this base pair substitution systematically occurs in homo- or heterozygous form in two non-transgenic varieties and seven transgenic varieties used to prepare CRMs. We analysed a larger number of varieties which potentially could be quantified in unknown test samples. Our results show that indeed the most frequently occurring allele in the tested collections is the one carrying the mismatch between

the forward primer and the DNA template. The allele matching the primer sequence was found in only a single individual in a total of 58 characterised varieties and lines (Table 2). The analysis of real-time PCR results revealed that this *Adh1 (70)* assay performs insufficiently. *Adh1 (70)* was identified as a negative outlier in terms of measured copy numbers in five varieties. This demonstrates that the *Adh1 (70)* measured copy number could vary in a large range. Interestingly, no attribution to variability in performance of real-time PCR could be assigned either to geographical origin of varieties or to the type of use of variety. These are the first experimental data and evidences on this matter, which certainly allow dispersing some existing doubts. In summary, the *Adh1 (70)* assay appears to be highly unreliable due to the frequent occurrence of the SNP in the primer attachment site resulting in unpredictable variation of the measured copy numbers. In contrast to this assay, *Adh1 (136)* RA demonstrated nucleotide stability in all tested MON 810 varieties and stable PCR performance in all tested conventional maize lines.

Currently, four RAs (*Hmg* and different *Adh1* assays) are part of official CRL validated methods. The methods for quantification of NK603, MON863 and GA21^{2005}, as well as a number of additional official methods for gene stacks, rely on the use of the *Adh1 (70)* assay [8] which was found in this study to be unreliable. Consequently, the reliability of these GMO quantification methods might be questioned.

With this study, we want to stress on the necessity to assess the uniformity of any newly developed RA prior to method validation. The occurrence of SNP variants in the genomic regions used to design a taxon or species RA and the frequency of SNP alleles, if occurring along suitable inbred lines, should be assessed *in silico* during the method development. In cases where this is not possible e.g. no sequence data available, the occurrence of SNP alleles has to be experimentally assessed in varieties/lines which might be used in GMO analysis. Two groups of such varieties could be distinguished: (1) varieties used to produce CRMs, also representing the varieties which would occur in the unknown samples, and (2) transgenic varieties containing the transgenic event which is targeted and one or more conventional varieties. The development of RAs for GMO quantification would significantly benefit from a harmonised requirement for testing the stable and uniform amplification along a wide range from 20 [8] to several hundreds of varieties [33]. The set of varieties selected to validate the uniformity of RA should represent the existing variation of the maize taxon at a certain moment. These criteria could be extended to other plant taxa relevant for GMO analysis as well.

References

1. James C (2009) Global Status of Commercialized Biotech/GM Crops in 2008. ISAAA Brief 39: Executive Summary
2. Directive 2001/18/EC of the European Parliament and the Council of 12 March 2001 on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC. Off J Eur Union L106:1–38
3. Regulation (EC) No. 1829/2003 of the European Parliament and the Council of 22 September 2003 on genetically modified food and feed. Off J Eur Union L266:1–23
4. Regulation (EC) No. 1830/2003 of the European Parliament and the Council of 22 September 2003 concerning the traceability and labelling of genetically modified organisms and the traceability of food and feed products produced from genetically modified organisms and amending Directive 2001/18/EC. Off J Eur Union L266:24–28
5. Commission Regulation (EC) No 641/2004 of 6 April 2004 on detailed rules for the implementation of Regulation (EC) No 1829/2003 of the European Parliament and of the Council as regards the application for the authorisation of new genetically modified food and feed, the notification of existing products and adventitious or technically unavoidable presence of genetically modified material which has benefited from a favourable risk evaluation. Off J Eur Union L102:14–25
6. Commission Recommendation 2004/787/EC of 4 October 2004 on technical guidance for sampling and detection of genetically modified organisms and material produced from genetically modified organisms as or in products in the context of Regulation (EC) No 1830/2003. Off J Eur Union L348:18–26
7. Commission Regulation (EC) No 1981/2006 of 22 December 2006 on detailed rules for the implementation of Article 32 of Regulation (EC) No 1829/2003 of the European Parliament and of the Council as regards the community reference laboratory for genetically modified organisms. Off J Eur Union L386:99–109
8. Community Reference Laboratory official website: http://www.crl-gmo.europa.eu
9. OECD (2003) Consensus document on the biology of *Zea mays* subsp. Mays (maize). Series on harmonization of regulatory oversight in biotechnology No.27. JT00147699
10. European Commission (2006) The common catalogue of varieties of agricultural plant species. Off J Eur Union C 68A/01
11. Hernandez M, Duplan M, Berthier G, Vaïtilingom M, Hauser W, Freyer R, Pla M, Bertheau Y (2004) J Agric Food Chem 52:4632
12. Taverniers I (2005) Development and implementation of strategies for GMO quantification in an evolving European context. PhD thesis. University of Ghent, ISBN 90-5989-049-3, p 346
13. Kuribara H, Shindo Y, Matsuoka T, Takubo K, Futo S, Aoki N, Hirao T, Akiyama H, Goda Y, Toyoda M, Hino A (2002) J AOAC Int 85:1077

14. Shindo Y, Kuribara H, Matsuoka T, Futo S, Sawada C, Shono J, Akiyama H, Goda Y, Toyoda M, Hino A (2002) J AOAC Int 85:1119

15. Yoshimura T, Kuribara H, Matsuoka T, Kodama T, Iida M, Watanabe T, Akiyama H, Maitani T, Furui S, Hino A (2005) J Agric Food Chem 53:2052

16. Brunner S, Fengler K, Morgante M, Tingey S, Rafalski A (2005) The Plant Cell 17:343

17. Emrich SJ, Li L, Wen T, Yandeau-Nelson M, Fu Y, Guo L, Chou H, Aluru S, Ashlock D, Schnable P (2007) Genetics 175:429

18. Liu K, Goodman M, Muse S, Smith JS, Buckler E, Doebley J (2003) Genetics 165:2117

19. Bhattramakki D, Dolan M, Hanafey M, Wineland R, Vaske D, Register J III, Tingey S, Rafalski A (2002) Plant Mol Biol 48:539

20. Tenaillon M, Sawkins M, Anderson L, Stack S, Doebley J, Gaut B (2002) Genetics 162:1401

21. Broothaerts W, Corbisier P, Schimmel H, Trapmann S, Vincent S, Emons H (2008) J Agric Food Chem 56:8825

22. Bru D, Martin-Laurent D, Philippot L (2008) Appl Environ Microbiol 74:1660

23. Ghedira R, Papazova N, Vuylsteke M, Ruttink T, Taverniers I, De Loose M (2009) Assessment of primer/template mismatch effects on real-time PCR amplification of target taxa for GMO quantification. J Agric Food Chem 57(20):9370

24. Zhang D, Corlet A, Fouilloux S (2008) Transgenic Res 17:393

25. Rozen S, Skaletsky HJ (2000) In: Krawetz S, Misener S (eds) Molecular biology. Humana Press, Totowa, p 365

26. Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208

27. Vandesompele J, De Preter K, Pattyn P, Poppe B, Van Roy N, De Paepe A, Speleman F (2002) Genome Biol 3: research0034.1–0034.11. Available from http://genomebiology.com/2002/3/7/research/0034.1

28. Papazova N, Ghedira R, Van Glabeke S, Bategi A, Windels P, Taverniers I, Roldan-Ruiz I, Van Bockstaele E, Milcamps A, Van den Eede G, Depicker A, De Loose M (2008) Plant Cell Rep 27:749

29. Sachs M, Dennis E, Gerlach W, Peacock W (1986) Genetics 113:449

30. Ceddia M, Rodriguez-Cerezo E (2008) A descriptive analysis of conventional, organic and GM crop and certified seed production in the EU. EUR 23443 EN

31. Vigouroux Y, Mitchell S, Matsuoka Y, Hamblin M, Kresovich S, Smith JSC, Jaqueth J, Smith OS, Doebley J (2005) Genetics 169:1617

32. Moor D (2005) Reliability of the Maize and Soybean Endogenous Reference Genes used in GMO Analysis if Food. PhD thesis, University of Bern, p 128

33. Chaouachi M, Giancola S, Romaniuk M, Laval V, Bertheau Y, Brunel D (2007) J Agric Food Chem 55:8003