JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

A Strategy for Designing Multi-Taxa Specific Reference Gene Systems. Example of Application––ppi Phosphofructokinase (ppi-PPF) Used for the Detection and Quantification of Three Taxa: Maize (*Zea mays*), Cotton (*Gossypium hirsutum*) and Rice (*Oryza sativa*)

MAHER CHAOUACHI,[†,‡] SANDRA GIANCOLA,[†] MARCEL ROMANIUK,[‡] VALÉRIE LAVAL,[‡] YVES BERTHEAU,[‡] AND DOMINIQUE BRUNEL*[,†]

Unité Etude du Polymorphisme des Génomes Végétaux (EPGV), Institut National de la Recherche Agronomique (INRA), Centre National de Génotypage (CNG), 2 Rue Gaston Crémieux 91057, CP5721, Evry Cedex, France, and Laboratoire de Méthodologies de la Détection des OGM (MDO), Institut National de la Recherche Agronomique (INRA), UR 256 PMDV/MDO, Route de Saint-Cyr, 78026 Versailles Cedex, France

In the first part of the paper, we report the description of a new strategy for the development of a plant reference gene system that can be used for genetically modified organism (GMO) analysis. On the basis of in silico research for candidate genes, the design of degenerate primers allowed the obtention of genomic sequences of the selected gene *ppi-phosphofructokinase* (*ppi-PPF*) for nine taxa in which GMOs have been developed. The comparison and the analysis of inter- and intraspecies sequence variability were performed using a large number of species and cultivars. As an example of application following the detection of single nucleotide polymorphism, we designed specific conventional and real-time polymerase chain reaction tests for the detection and quantification of three taxa, namely, maize, cotton, and rice. This system was highly specific and sensitive. The gene copy number conservation among different cultivars was analyzed and confirmed with a sequencing step. This reference gene system is adequate for use in routine assays for the quantification of GMOs. We then explain briefly the constraints faced and propose recommendations when designing a reference gene system depending on the species to be targeted.

KEYWORDS: Rice; cotton; maize; real-time PCR; *ppi-phosphofructokinase*; quantification; GMO; bottlenecks

INTRODUCTION

According to the ISAAA (International Service for the Acquisition of Agri-biotech Applications) (www.isaaa.org), the area of GMO (genetically modified organism) cultivation worldwide increased in 2006 by 12% as compared to the year 2005. Soybean continued to be the principal GMO crop in 2006, occupying 58.6 million ha, followed by maize (25.2 million ha), cotton (13.4 million ha), and canola (4.8 million ha). Worldwide, there exists a large panoply of genetically modified plants from different families such as *Brassicaceae* (e.g., rapeseed), *Fabaceae* (e.g., soybean), *Chenopodiaceae* (e.g., sugar beet), or *Solanaceae* (e.g., potato).

Currently, more than 18 GMO crops are approved in the European Union (EU) corresponding to different species and conferring a variety of agronomic traits. In this context, an effort is underway in the laboratories of the European Commission (EC) and the member states to develop molecular approaches for the detection and quantification of GMOs in food and feed products containing ingredients from GMOs. There are mandatory rules for the labeling of all GMO foods or food ingredients: For example, in the EU, the threshold is 0.9% (*1*) in Japan; 5% in Taiwan and South Korea (*2*), and 4% in Brazil (*3*). During the last 4 years, many European GMO research projects have been established (*4*) to provide appropriate methods to detect and quantify GMOs to comply with the European regulations.

The availability of reliable and precise methods and tests for GMO quantification and traceability is essential to distinguish between genetically modified and conventional foods. DNA is widely used as the target for GMO analysis because of its

* To whom correspondence should be addressed. Tel: +33-01-60-87-83-31. Fax: +33-01-60-87 83-83. E-mail: brunel@versailles.inra.fr.
[†] Unité Etude du Polymorphisme des Génomes Végétaux (EPGV).
[‡] Laboratoire de Méthodologies de la Détection des OGM (MDO).

10.1021/jf071429d CCC: $37.00
Published on Web 09/08/2007

Table 1. Summary of Reference Genes Found in the Bibliography

species	target(s)	accession no.	refs
Zea mays (maize)	*Adh1* (alcohol dehydrogenase 1)	X04050	*17*
	Hmga (high mobility group protein)	AF127919	
	Ivr1 (invertase)	U16123	
	Zein	X07535	
Brassica napus (rapeseed)	*Cruciferin*	X59294	*7, 9*
	BnACCg8 (acetyl CoA carboxylase)	X77576	*7, 8*
	Hmg I/Y (high mobility group protein)	AF127919	*10*
Solanum tuberosum (potato)	*UGPase* [uridine diphosphate (UDP)-glucose pyrophosphorylase]	U20345	*11, 12*
	Pci (metallo-carboxypeptidase inhibitor)	AF060551	*13*
Oryza sativa (rice)	*SPS* (sucrose-6-phosphate synthtase)	U33175	*14*
	Oryzain β	D90407	*15*
	Gos 9	X51909	*15*
Solanum lycopersicum (tomato)	*Mcpi* (metallo-carboxypeptidase inhibitor)	X59282	*13*
	Lat52 (putative proteine 18 kDa)	19263	*16*
Hordeum vulgare (barley)	γ*-hordein*	M36378	*15*
	Pkaba1(serine/threonine protein kinase)	AB058924	*6*
	Acc1 (acetyl CoA carboxylase1)	AF029895	*15*
Triticum aestivum (wheat)	*RALyase*	AB032124	*15*
	Wax-D1	AF113844	*18*
	Pkaba1 (serine/threonine protein kinase)	M94726	*6*
Gossypium hirsutum (cotton)	*Sad1* (stearoyl-ACP desaturase)	AJ132636	*19, 20*
Glycine max (soybean)	*Lectin*	K00821	*15*
Helianthus annuus (sunflower)	*Helhianthin* (11s storage protein)	M28832	*15*

stability and high detectability in processed matrixes. The method of choice used routinely in most analytical laboratories for GMO detection is polymerase chain reaction (PCR), including real-time PCR techniques. In practice, the GMO percentage is calculated as a ratio of the specific GMO target (*5*) with respect to an endogenous plant gene (reference gene). It is for this reason that a sensitive and specific method of reference gene detection, which ideally can be applied to multiple taxa, is essential.

Although many research groups have previously developed specific tests to be used as reference genes in various crops, few are validated through interlaboratory studies and nearly none can be applied to multiple taxa. A reference gene must be nuclear, taxon specific, present in low copy number, and have low intraspecies variability. **Table 1** lists the reference genes developed in different crops. For the development of all of the tests cited in **Table 1**, performance criteria such as sensitivity and specificity were applied to validate their use as endogenous genes for GMO quantification. The design of these tests is always based on database searches, for example, the National Centre for Biotechnology Information (NCBI), and in general, one gene was used for the quantification of one species. The *PKABA1* sequence (*6*) can discriminate two species (wheat and barley) using one couple of consensus primers; however, the wheat probe did not differentiate between *Triticum aestivum* and *Triticum durum*. Additionally, in a previous European program "GMOchips" (2001–2004), two reference genes, *RuBisco activase* (rbc) and *sucrose synthase* (ss), were used for the detection of GMOs with microarrays. These genes allowed the recognition of five plant species (maize, soybean, rapeseed, tomato, and sugar beet) (*3*) but did not allow the differentiation of the 21 other cited species. To date, no specific tests exist that discriminate more than two taxa in which GMOs have been developed. Therefore, a specific multitaxa reference gene system will allow the establishment of common reference gene tests between species and between different laboratories.

Sequencing is essential for the detection of inter- or intraspecies nucleic acid variability. This has not been done in the previously developed reference genes. Furthermore, the number of cultivars previously used is insufficient for the validation of the specificity. Ideally, it is necessary to test a number of representative cultivars, including wild cultivars and related species, chosen in function of taxonomy, phylogenetic relatedness, and ploidy.

Here, we describe a new strategy for accurate quantification of endogenous genes based on sequencing and the design of degenerate and specific primers followed by real-time PCR with TaqMan chemistry. The aim was to find one region in the sequence of a gene that allows us to differentiate a maximum of taxa in which GMOs have been developed. This will eliminate the future need to search for a reference gene for each new species that is subjected to GMO development. As an example of application, one selected region of the *ppi-phosphofructokinase* (*ppi-PPF*) gene was sequenced in nine different species using degenerate primers. After that, primers and probes were designed for the development of three highly sensitive and specific tests to differentiate three species (maize, cotton, and rice).

Furthermore, in a second part of the paper, we point out the bottlenecks faced when developing a reference gene system. We discuss additional criteria that should be added to the list of classic requirements for the selection of an endogenous gene as a reference gene for GM DNA quantification (species specific, low copy number, nuclear, etc.).

MATERIALS AND METHODS

Plant Materials. A collection of 365 cultivars of nongenetically modified maize (*Zea mays*), one accession of *Arabidopsis thaliana*, one cultivar of alfalfa (*Medicago truncatula*), one line of pea (*Pisum sativum*), 57 wheat lines (46 *T. aestivum* and 11 *T. durum*), one line of cauliflower (*Brassica oleraceae*), one chinese cabbage (*Brassica rapa*), one line of mustard (*Brassica juncea*), and one line of black mustard (*Brassica nigra*) were provided by different INRA groups (Versailles, Rennes, Avignon, and Clermont Ferrand). One line of sugar beet (*Beta vulgaris*), one line of rapeseed (*Brassica napus*), 18 lines of cotton (17 *Gossypium hirsutum* and one *Gossypium barbadense*), 20 lines of rice (*Oryza sativa*), one line of white bean (*Phaseolus aureus*), one line of cereal rye (*Secale cereale*), and the ancestor of maize, teosinte (*Zea diploperennis*), were cultivated in greenhouses at the Institut National de la Recherche Agronomique (INRA, Versailles, France). All plant material was provided as leaves.

Isolation and Quantification of Plant Genomic DNA. For the isolation of DNA, 1 g of leaves was ground in liquid nitrogen. The

Table 2. Sequences of Primers and Probes Used for the Qualitative and Quantitative PCR Tests

primer/probe name	orientation	sequence (5′–3′)	amplicon size	ref
Degex5F	sense	GACAAGATTGARACNCCNGA	cotton (450 bp), maize (500 bp), and rice (719 bp)	this study
Degex5R	antisense	CAAGCAGTGTCRAANCCRAA		
CotPr	probe	FAM-CAGCAGCAAAGCTTGATCTGGATGG-MGB		this study
CotF	sense	AGAGTTGGTACGATTTTCAGTTCAAG	262 bp	
CotR	antisense	TGGGCATCCAATCACTCGAG		
SAD1Pr	probe	HEX-TCACCCACTCCATGCCGCCTCACA-TAMRA		*36*
S1F	sense	CCAAAGGAGGTGCCTGTTCA	107 bp	
S1R	antisense	TTGAGGTGAGTCAGAATGTTGTTC		
RiPr	probe	FAM-ACACTGTAAACAAAC-MGB		this study
RiF	sense	AATTCTGTCATGTATTTGAGCAGTTCA	79 bp	
RiR	antisense	AATGACAACAAGCCCATCCAA		
SPSPr	probe	HEX-GACGCACGGACGGCTCGGA-MGB		*15*
SPSF	sense	TTGCGCCTGAACGGATAT	81 bp	
SPSR	antisense	CGGTTGATCTTTTCGGGATG		
MaiPr	probe	VIC-CAGCCAACAAGCT-MGB	148 bp	this study
MaiF	sense	TGTTGTGCAGTTCAAGCAAGCT		
MaiR	antisense	TGGAACACCAAAACACAGTGTCA		
Adh1-Pr	probe	FAM-AATCAGGGCTCATTTTCTCGCTCCTCA-TAMRA		*7*
Adh-F3	sense	CGTCGTTTCCCATCTCTTCCTCC	136 bp	
Adh-R4	antisense	CCACTCCGAGACCCTCAGTC		

Table 3. Description of the PCR Programs Used in This Study

PCR programs							
"touch down" with degenerate primers		qualitative PCR		real-time PCR		sequencing	
94 °C (4 min)		94 °C (4 min)		94 °C (10 min)		94 °C (5 min)	
94 °C (1 min) 65 °C (1 min) until 55 °C 72 °C (2 min)	10 cycles	94 °C (30s) 63 °C (30 s) (maize and rice) 60 °C (cotton) 72 °C (30 s)	40 cycles	92 °C (15 s) 63 °C (1 min) (maize and rice) 60 °C (cotton) 72 °C (30 s)	40 cycles	94 °C (10 s) 53 °C (5 s) 60 °C (4 min)	26 cycles
94 °C (45 s) 55 °C (45 s) 72 °C (1 min)	30 cycles						
72 °C (6 min)		72 °C (7 min)		15 °C (5min)		15 °C (∞)	

cetyltrimethylammonium bromide (CTAB) DNA extraction protocol used is described in prEN ISO/DIS 21571 (*21*). The DNA quantities were estimated using the PicoGreen dsDNA quantification kit (Invitrogen France, Cergy Pontoise, France) with a SpectraMAX Gemini spectrofluorometer (Molecular Devices, St. Grégoire, France). Standards were five different dilutions of calf thymus DNA (0–50 ng/μL). Another rapid DNA extraction protocol (*22*) was tested for the extraction of *B. nigra*, *B. juncea*, and *T. durum*. The quality of DNA in PCR reactions was assessed using universal primers (*23*) (data not shown).

TaqMan Probe and Primer Design. All of the primers and probes used in the present study are presented in **Table 2**. Primers for the extension reaction were determined using OLIGO version 6.0 and were purchased from MWG-Biotech AG (Ebensburg, Germany). The design of minor groove binder (MGB) probes for the TaqMan assay was carried out using the Primer Express software, version 2.0 (Applied Biosystems, Foster City, CA). The 3′-MGB was used for probes to discriminate single mismatches. All probes and primers were supplied by Applied Biosystems (Applied Biosystems).

PCR Conditions and Optimization (Qualitative PCR, QRT-PCR). Qualitative PCR was run on a MJ Research (Biorad, Waltham, United States) thermocycler; PCR amplifications were performed in a final volume of 25 μL. Each reaction mixture contained 1× PCR buffer, 0.2 mM dNTP, 0.4 μM each primer, 25 ng of each DNA sample, 1 unit of Platinum Taq DNA Polymerase (Invitrogen, United States), and 3% dimethyl sulfoxide (DMSO) for the rice test. To avoid nonspecific amplification, an optimization of the annealing temperature was necessary. For this, PCR parameters were varied, including $MgCl_2$ concentration, primer concentrations, DMSO concentration, and the annealing temperature. Thus, different conditions were used depending on the target and the primers (see **Table 3** for PCR programs). The amplicons were resolved by electrophoresis in 3% agarose gels (Invitrogen) run in Tris acetate EDTA buffer for 70 min at 110 V and stained with ethidium bromide (0.4 ng/mL) for 20 min. The gels were scanned using Image Master VDS (Amersham Pharmacia Biotech, Freiburg, Germany). Real-time PCR reactions were run on an ABI PRISM 7900 HT Sequence detection system (Applied Biosystems). Both TaqMan and SyberGreen reactions were performed in 5 μL final assay volume. SyberGreen reactions contained 2.4 μL of Power SYBR Green PCR Master Mix (Applied Biosystems) and 1 μM each of the specific primers. The TaqMan PCR final reaction contained 1 μM the primers and 200 nM TaqMan-MGB probes added to 2.36 μL of master mix [TaqMan Universal PCR Master Mix No Amperase UNG (2×)]. The DNA concentration used was 5 ng/μL.

DNA Sequencing. Primers for sequencing are shown in **Table 2**. The PCR products were purified using P100 (Biorad Bio-Gel R P-100 Gel Fine 45–90 μm, Hercules, CA). The sequencing reaction was performed in a 10 μL reaction containing 10 nmol/L of primer (forward or reverse), 1 μL of BigDye Terminator Cycle Sequencing reaction mixture (Perkin-Elmer), 1 μL of Big Dye buffer, 5 μL of H_2O, and 2 μL of the purified PCR product. The products of this reaction were purified using a G50 gel filtration (Sephadex TM G-50 superfine; Amersham Biosciences, Hercules, CA) and loaded onto an ABI3730XL 96 capillary sequencer. Sequence alignments and detection of polymorphism were performed using the software GENALYS, available at the website http://software.cng.fr.

RESULTS AND DISCUSSION

Description of the Strategy. Until now, the selection of reference genes was single taxon specific and no sequences were found to differentiate two or more species (*7*). **Figure 1** illustrates our three-part strategy described below.

Part 1: In Silico Search for the Candidate Gene. *Step 1. Selection of Suitable Candidate Genes.* For choosing reference genes, it is ideal to begin with a sequence already present in a

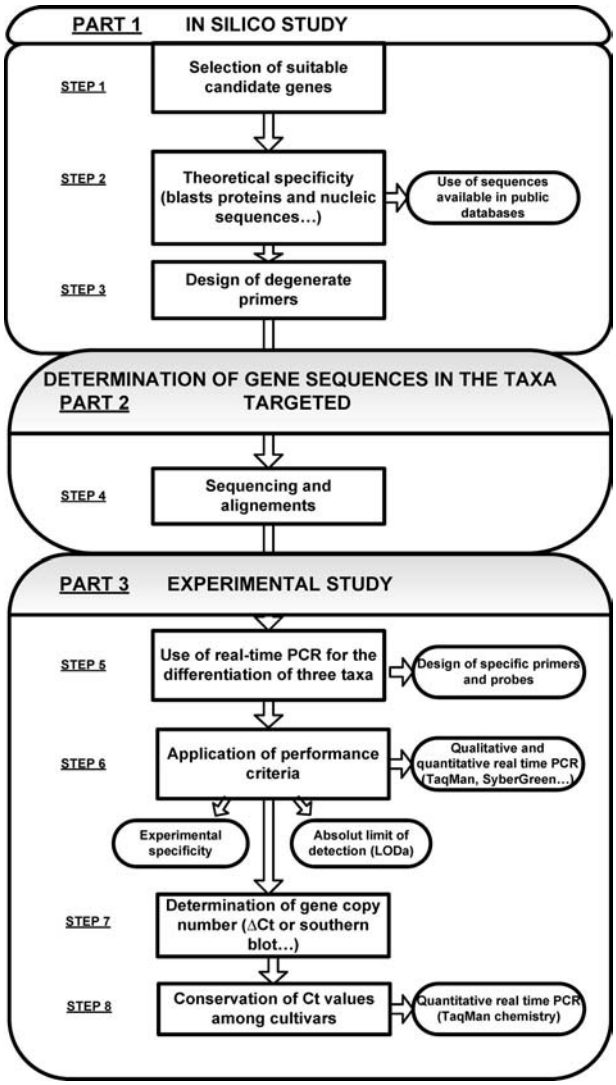

Figure 1. This diagram shows the different steps proposed in this publication for the design of a reference gene system. This strategy is divided into three parts and is composed of eight analytical steps.

known and fully sequenced genome. However, there is a lack of genomic information in many crop species and only cDNA or partial sequences are available. Fortunately, complete genomic sequences and many genetic maps are available for a few "model" plant species such as *Arabidopsis* and rice (*24*). Furthermore, many projects are underway to make more genomic sequences available and accessible to scientists in databases such as NCBI (http://www.ncbi.nlm.nih.gov/). Our objective was to find a sequence presenting enough interspecies polymorphism to differentiate nine targeted species (maize, rapeseed, soybean, tomato, potato, sugar beet, cotton, and rice). The database of Conserved Ortholog Set markers (COS) (http://www.sgn.cornell.edu/search/marker/cos_list.pl) was used. In this database, a set of highly conserved single copy genes, used as markers for comparative mapping between tomato and *Arabidopsis* genomes, are described. After the selection, the homology of the sequences was analyzed in the other species. The preselection began with 1028 unique genes in tomato.

Step 2. Study of the Theoretical Specificity. Each sequence was extracted from the COS database, blasted in NCBI to search for homologies, and finally aligned using the program "ClustalW" (http://www.genome.jp). These steps allowed us to collect a maximum of sequences from the species targeted. We also used The Adaptive Evolution Database (TAED) (http://www.bioinfo.no/tools/TAED) to search for conserved genes (*25*, *26*). This contains information from 4294 higher plants. After the selection of several candidate genes based on their metabolic function and the availability of their genomic sequences, the gene encoding the β-subunit of putative *ppi-PPF*, also known as pyrophosphate-fructose 6-phosphate 1-phosphotransferase, was chosen. This protein has two subunits, α and β, the latter which is our target in this study. It catalyzes the reversible conversion of fructose 6-phosphate and pyrophosphate to fructose 1,6-bisphosphate and inorganic phosphate (*27*).

Step 3. Design of the Degenerate Primers. Degenerate primers were designed using the full sequence of the gene (exon and intron) in the rice genome sequence database. First, four couples of degenerate primers targeting the exons 4–7 were designed and were tested on nine different species (tomato, potato, rice, cotton, soybean, cabbage, cauliflower, sugar beet, and maize). On the basis of the intensity of the band for all of the species, one couple of primers targeting exon 5 was retained. The sizes of the PCR products with degenerate primers were approximately 450 bp for cotton; 500 bp for soybean and maize; 550 bp for rapeseed, cabbage, and cauliflower; 600 bp for potato and tomato; 650 bp for sugar beet; and 719 bp for rice. The other primer pairs gave rise to weak bands and were not specific for some species (data not shown). After the PCR reaction, all of the PCR products were sequenced, and the sequences were aligned to detect the inter- and intraspecies polymorphism. The sequences obtained are available in Genbank (accession numbers EF445627, EF445628, EF445630, EF445631, and EF445632).

Part 2: Determination of Gene Sequences in the Taxa Targeted. *Step 4. Sequencing and Alignments.* After sequencing, the alignment of the sequences (Supporting Information, **Figure 1**) obtained showed a conserved region corresponding to exon 5 in all of the species tested and considerable polymorphism in the introns. This allowed the development of specific Taqman probes for endogenous reference gene detection, required for GMO quantification. As an example of application, three probes targeting maize, cotton, and rice were designed.

Part 3: Experimental Study of the Candidate Gene. *Step 5. Use of Real-Time PCR for the Differentiation of Three Taxa.* Nine species were tested in real-time PCR. Initially, three tests were successfully developed on maize, rice, and cotton. For several other species, certain constraints did not allow the development of a specific test (see below). **Figure 2** illustrates the position of degenerate and specific primers and probes developed in this study. Primers were designed in the introns or at the intron–exon border, to increase their specificity, whereas the probes were designed in the exons. A reliable PCR system must be able to amplify DNA from different plant species with the same efficiencies independently of the genotype from which it is derived. In conventional PCR, the constant band intensity and the size of the product indicated the absence of polymorphism. Nonetheless, sequencing of the amplicons from a large number of diverse cultivars corresponding to "core collections" is essential for the validation of the primers.

Step 6a. Specificity of the Primers and Probes. Specificity is one classical criterion used for the assessment of method performance. It is defined according to several international guidelines (*28*, *29*) to be the property of a method to respond exclusively to the characteristic or analyte of interest. In a first round, the specificity was tested by both qualitative and quantitative PCR (SyberGreen, data not shown) on a few species and varieties. The specificity was tested in conventional PCR using 25 ng of template DNA from all of the available varieties of the target species. The size of PCR amplicons was 148 bp

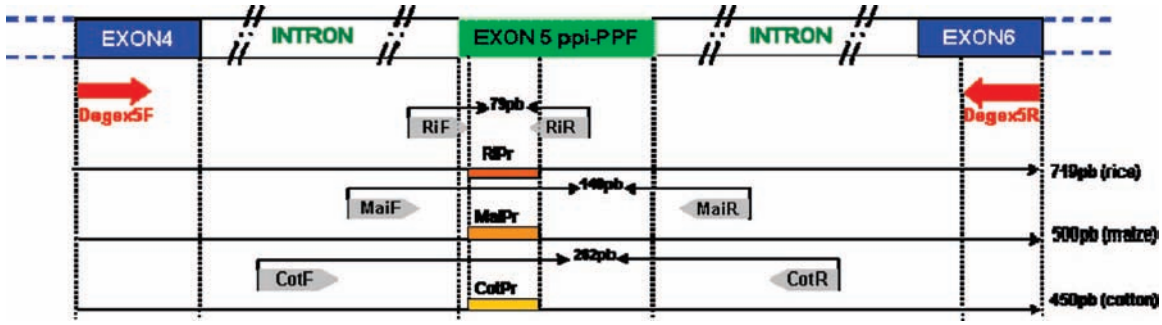

Figure 2. Schematic illustration of the targeted region of *ppi-PPF*. Positions of probes, degenerate and specific primers, and the respective amplified fragment in cotton, rice, and maize are indicated.

Table 4. Study of Specificity with the Described Plant Reference Gene System[a]

	ppi-PPF		
species	maize probe	rice probe	cotton probe
Arabidopsis thaliana acc. *(Col-o)*	−	−	−
Beta vulgaris var. *belluga*	−	−	−
Brassica napus var. *westar*	−	−	−
Brassica oleracea var. *HDEM*	−	−	−
Brassica carinata var. *awassa19*	−	−	−
Brassica juncea var. *junius*	−	−	−
Brassica rapa var. *Z1*	−	−	−
Capsicum annuum var. *fushimi*	−	−	−
Glycine max var. *safrana*	−	−	−
Gossypium barbadense	−	−	+
Gossypium hirsutum[b]	−	−	+
Hordeum vulgare var. *merlot*	−	−	−
Linum usitatissimum var.*humile*	−	−	−
Solanum lycopersicum var. *nagcarlan*	−	−	−
Medicago trunculata var. *SA009707*	−	−	−
Orya sativa[b]	−	+	−
Phaseolus mungo var. *aureus*	−	−	−
Pisum sativum var. *champagne*	−	−	−
Secale cereale var. *multicaule*	−	−	−
Solanum tuberosum var. *operle*	−	−	−
T. durum[b]	−	−	−
Triticum aetivum[b]	−	−	−
Zea diploperennis"teosinte"[c]	+	−	−
Zea mays[b]	+	−	−

[a] End point PCR with *ppi-PPF* primers was performed on DNA extracted from different plant species and varieties. (−) means that no amplification was observed. (+) means that a band corresponding to the correct size was observed. [b] Names of the varieties are listed in the Supporting Information tables. [c] Note that "teosinte" being the ancestor of maize gives a positive signal as expected.

for maize, 261 bp for cotton, and 79 bp for rice with the *ppi-PPF* primers. The electrophoretic analysis of qualitative PCR revealed that no aspecific amplification products were observed for any of the species tested, while specific bands at the expected size were found for target species and varieties. **Table 4** resumes the results of the PCR reactions. Specificity was also tested in real-time PCR, and no fluorescent signals were detected with species other than the expected ones.

Step 6b. Sensitivity of the System. Sensitivity is another routine criterion used for the assessment of method performance. The parameter used for the estimation of the sensitivity is the absolute limit of detection (LODa). It is the lowest amount or concentration of analyte in a sample that can be reliably detected but not necessarily quantified. To determine the LODa, a dilution series containing from 20000 to 0.2 HGC (haploid genome copy) was analyzed in triplicate in conventional and real-time PCR assays. HGC or copy numbers of the target gene (*ppi-PPF*) in genomic DNA were derived from the haploid genomic weight of 2.6 pg for maize, 2.33 pg for cotton, and 0.89 pg for rice (*30*). As expected, the ability to detect the targeted species decreased with decreasing copy numbers. In our case, the LODa was 2 HGC for rice and maize and 20 HGC for cotton. The linear correlation coefficient (R^2) of *ppi-PPF* ranged from 0.95 to 0.99. On the basis of the slopes of the standard curves, the amplification efficiencies ranged from 92 to 99%, as calculated with the formula: $E = [10^{(-1/\text{slope})} - 1]$. The high efficiency and correlation coefficients make the test developed in the present study suitable for calculating reference gene quantities.

Step 7. Estimation of the Copy Number of the Gene in the Different Species. A reference gene, to be classified as such, should be low copy number. We tried to confirm this criterion in the three taxa targeted. Classically, gene copy numbers are estimated by Southern analysis, which is laborious and time consuming and requires relatively large amounts of plant material (*31*). Real-time PCR has been used frequently to estimate the copy number of transgenes (*32*, *33*). Studies have shown a strong correlation between Southern blot and real-time analysis (*34*). We calculated the copy number of *ppi-PPF* by comparing the quantitative PCR results with those of publicly available endogenous reference genes with known copy numbers, namely, the gene *sad1* for cotton (2 HGC) (*19*), *sps* for rice (1 HGC) (*14*), and *adh1* for maize (1 HGC) (*17*). We also compared the amplification efficiency in the same PCR conditions, using the same samples and the same DNA dilution series. We found similar efficiencies in all of the genes tested, and the calculation of copy number (based on ΔC_t values) demonstrated appropriate values for all three species.

Step 8. Intraspecies Conservation of Gene Copy Number. Another required condition when designing a new reference gene is to have a constant copy number across different cultivars of the target species. All cultivars of all of the species were analyzed by conventional PCR, and the expected band sizes were obtained. Nonetheless, the only way to check for sequence variation is by sequencing. Sequencing of all amplicons confirmed our previous results. No allelic variation was observed for the *ppi-PPF* sequence chosen for the 365 varieties of maize, 18 varieties of cotton, and 20 varieties of rice tested. The absence of variation among cultivars confirmed the low copy number of the target gene (*34*). Moreover, similar C_t values were obtained for rice (27.127 ± 0.43), cotton (29.59 ± 0.32) and maize (26.43 ± 0.25) when real-time PCR tests were performed on DNA of all of the cultivars for each species. This conservation of the C_t values enforces the idea that the copy number of this gene among cultivars is stable. It should also be noted that no significant differences of C_t values were detected between wild and cultivated species. The low standard deviations (SD) observed in C_t values are probably due to the high quality of extracted DNA and the means of quantification (*12*).

Bottlenecks When Designing a Specific Plant Reference Gene System. Three main problems for the determination of a *ppi-PPF* specific reference system for other species (rapeseed,

potato, tomato, and sugar beet) were encountered as follows: (i) differentiation of phylogenetically related species, (ii) introgressions that can occur between cultivated and wild accessions belonging to the same family, and (iii) the ploidy level. Below, we discuss these problems further, as well as potential solutions.

Case 1. Phylogenetically Related Species: Example of Potato (*Solanum tuberosum*) and Tomato (*Solanum lycopersicum*). The strategy described above was applied to three species (maize, cotton, and rice) that have no close phylogenetic relations. What happens when designing a reference gene system for related species such as tomato (*S. lycopersicum*) and potato (*S. tuberosum*)?

Much work has contributed significantly to the understanding of the relationships of tomato and potato and has resulted in a phylogenetic classification of the Solanaceae using data from the chloroplast genome (*35*). The data demonstrated that these genera have highly conserved genomes.

Our purpose is to search for a sequence allowing the differentiation between potato and tomato. Because the two species are phylogenetically close, a strong similarity at the DNA level was observed in exon 5 of the *ppi-PPF* gene employed in this paper. The SNPs (single nucleotide polymorphism) observed in positions 25 and 39 of exon 5 and confirmed by sequencing (accession number M55191 for potato and SGN-U214522L in the Cornell database for tomato) did not allow the discrimination of the two species (data not shown). To this end, more regions of the gene need to be explored to find enough differences for the design of specific primers.

In addition, a representative collection of samples must be tested for the validation of specificity. It should include not only related species such as eggplant (*Solanum melongena*), pepper (*Capsicum annuum*), *Solanum aethiopicum* and *Solanum chacoense* but also wild accessions such as *Solanum pennellii* (*36*).

Case 2. Occurrence of Introgression during Plant Breeding: Example of Sugar Beet (*B. vulgaris*). The sequence of *ppi-PPF* obtained with degenerate primers did not allow the differentiation between wild and cultivated sugar beets. However, it can be used for genus discrimination. Taxonomically, sugar beet belongs to the genus *B. vulgaris* ssp. *vulgaris* L. within the Chenopodiaceae family. This genus includes four genomic sections or genepools (Beta, Corrolinae, Nanae, and Procumbentes) (*37*). It is known that sugar beet is a wind-pollinated out-crossing crop. Thus, the occurrence of introgression (especially of the resistance genes to rhizomania) is frequent. Sugar beet is sexually compatible with many wild species from the section Beta but less compatible with species from the three other sections (*38*). The development of a specific reference gene for sugar beet must discriminate among the cultivated beets classified in the same Beta section (fodder beet, garden beet, and leaf beet) on one hand and from the other genepools on the other hand. To date, the only published report describing specific primers for sugar beet GMO quantification is the CRL-validated method (http://gmo-crl.jrc.it/) using the g *lutamine synthase gene 2* for the quantification of transgenic sugar beet (event T7-1) (http://www.agbios.com). The development of more widely applicable reference gene systems requires the establishment of a representative collection of samples including all genepools and a large number of cultivated sugar beets. Subsequently, the selection of a nuclear gene, followed by an experimental validation step using sequencing and the localization of conserved regions, will allow the design of a specific test that can be performed using a core collection.

Case 3. Influence of Ploidy Level: Example of Rapeseed (*B. napus*). We tried to extrapolate the results found with *ppi-PPF* for the detection of rapeseed (*B. napus*), but the results showed a lack of primer specificity between the *Brassica* crops that are described in the triangle of U (*39*). In fact, the same PCR product was found not only with *B. napus* (AACC, $2n = 38$) but also with *B. juncea* (mustard; AABB, $2n = 36$), *B. oleracea* (cauliflower; CC, $2n = 18$), and *B. rapa* (chinese cabbage; AA, $2n = 20$). Nonetheless, sequencing revealed SNPs between these species (Supporting Information, Figure 1). This means that the gene was present in both genomes (AA and CC). Rapeseed (*B. napus*) is an amphidiploid species and is the result of a cross between two parental diploid genomes: *B. oleracea* and *B. rapa*. We tested publicly available genes specific to *B. napus*, such as *cruciferin* (*9*), *acetyl-CoA carboxylase* (*8*), and *hmg I/Y* (*10*), but amplification of other *Brassicaceae* species was detected (data not shown). This lack of specificity was also confirmed in other studies (*40*). Because rapeseed results from the addition of two different genomes, it should be possible in the future to identify SNPs, which distinguish parental genotypes vs the *B. napus*genotype and to design the corresponding probes.

Conclusion. We have described a new strategy for developing a multitaxa specific reference gene system based on SNPs and the design of TaqMan probes. In silico research of candidate genes using different databases was performed. Primers and probes developed on the chosen *ppi-PPF* gene allowed the detection and the quantification of species specific DNA in maize, rice, and cotton. This specificity was higher than the hybridization of probes on chips previously described. These primers and probes may be used in routine analysis after interlaboratory validation. The designed degenerate primers will allow other laboratories to develop specific TaqMan tests based on the sequence of *ppi-PPF* in their target species in less time.

We have used the collection of plants for the specificity tests. These collections were composed of common cultivated species, wild accessions, and related species. Our work stresses that a thorough knowledge of a given species is required to allow a well-informed choice of the material to be included in developing a reference gene system and thus avoids problems of nonspecificity.

During the experimental part of the present study, a lack of specificity was observed when using the *ppi-PPF* sequence for the detection of some species such as tomato, potato, sugar beet, and rapeseed. We have pointed out and discussed three potential constraints when designing reference gene systems in relation to these three species, namely, use of phylogenetically close species (tomato and potato), species involved in the introgression of genes (sugar beet), and species composed of more than one genome (rapeseed).

ACKNOWLEDGMENT

We thank all INRA groups for providing us plant and DNA samples. We thank Heather McKhann for her helpful comments and corrections on the manuscript.

Supporting Information Available: Collection of lines (name and origins) of the maize, cotton, rice, and wheat taxa (Table 1, maize collection; Table 2, cotton collection; Table 3, rice collection; and Tables 4 and 5, wheat collection). An alignment of the region targeted in nine taxa (rice, maize, cotton, soybean, sugar beet, tomato, potato, cabbage, and cauliflower) (Figure 1). This material is available free of charge via the Internet at http://pubs.acs.org.

LITERATURE CITED

(1) Regulation (EC) No. 1830/2003 of the European Parliament and of the Council of 22 September 2003 concerning the traceability and labeling 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. Commun.: Legis.* 2003, *L268*, 24–28.

(2) Hino, A. Safety assessment and public concerns for genetically modified food products: The Japenese experience. *Toxicol. Pathol.* **2002**, *30*, 126–128.

(3) Leimanis, S.; Hernandez, M.; Fernandez, S.; Boyer, F.; Burns, M.; Bruderer, S.; Glouden, T.; Harris, N.; Kaeppeli, O.; Philipp, P.; Pla, M.; Puidomenech, P.; Vaitilingom, M.; Bertheau, Y.; Remacle, J. A microarray-based detection system for genetically modified (GM) food ingredients. *Plant Mol. Biol.* **2006**, *61*, 123–139.

(4) Kubista, M.; Andrade, J. M.; Bengtsson, M.; Forooton, A.; Jiri, J.; Lind, K.; Sindelka, R.; Sjöback, R.; Björn, S.; Stömbom, L.; Stahlberg, A.; Zoric, N. The real-time polymerase chain reaction. *Mol. Asp. Med* **2006**, *27*, 95–125.

(5) Fernandez, S.; Charles-Delobel, C.; Geldreich, A.; Berthier, G.; Boyer, F.; Collonnier, C.; Coue-Philippe, G.; Diolez, A.; Duplan, M. N.; Kebdani, N.; Romaniuk, M.; Feinberg, M.; Bertheau, Y. Quantification of the 35S promoter in DNA extracts from genetically modified organisms using real-time polymerase chain reaction and specificity assessment on various genetically modified organisms, part I: operating procedure. *J. AOAC Int.* **2005**, *2*, 547–557.

(6) Ronning, S. B.; Knut, G. B.; Boydler Andersen, C.; Holst-Jensen, A. Novel reference gene, PKABA1, used in a duplex real-time polymerase chain reaction for detection and quantitation of wheat- and barley-derived DNA. *J. Agric. Food Chem.* **2006**, *54*, 682–687.

(7) James, D.; Schmidt, A. M.; Wall, E.; Green, M.; Masri, S. Reliable detection and identification of genetically modified maize, soybean and canola by multiplex PCR analysis. *J. Agric. Food Chem.* **2003**, *51*, 5839–5834.

(8) Hernández, M.; Río, A.; Esteve, T.; Prat, S.; Pla, M. A rapeseed-specific gene, acetyl-CoA carboxylase, can be used as a reference for qualitative and real-time quantitative PCR detection of transgenes from mixed food samples. *J. Agric. Food Chem.* **2001**, *49*, 3622–3627.

(9) Block, A.; Schwarz, G. Validation of different genomic and cloned DNA calibration standards for construct-specific quantification of LibertyLink in rapeseed by real-time PCR. *Eur. Food Res. Technol.* **2003**, *216*, 421–427.

(10) Weng, H.; Yang, L.; Liu, Z.; Ding, J.; Pan, A.; Zhang, D. Novel reference gene, High-mobility-group protein I/Y, used in qualitative and real-time quantitative polymerase chain reaction detection of transgenic rapeseed cultivars. *J. AOAC Int.* **2005**, *88*, 577–584.

(11) Watanabe, T.; Kuribara, H.; Mishima, T.; Kituchi, H.; Kodama, T.; Futo, S.; Kasama, K.; Toyota, A.; Nouno, M.; Saita, A.; Takahashi, K.; Hino, A.; Akiyama, H.; Maitani, T.; Kubo, M. New qualitative detection methods of genetically modified potatoes. *Biol. Pharm. Bull.* **2004**, *27*, 1333–1339.

(12) Rho, J. K.; Lee, T.; Jung, S. I.; Kim, T. S.; Park, Y. H.; Kim, Y. M. Qualitative and quantitative PCR methods for detection of three lines of genetically modified potatoes. *J. Agric. Food Chem.* **2004**, *52*, 3269–3274.

(13) Hernandez, M.; Ferrando, A.; Esteve, T.; Puigdomenech, P.; Prat, S.; Pla, M. Real-time and conventional polymerase chain reaction systems based on the metallo-carboxypeptidase inhibitor gene for specific detection and quantification of potato and tomato in processed food. *J. Food Prot.* **2003**, *66*, 1063–1070.

(14) Ding, J.; Jia, J.; Yang, L.; Wen, H.; Zhang, C.; Liu, W.; Zhang, D. Validation of a rice-specific gene, *sucrose-phosphate synthase*, used as the endogenous reference gene for qualitative and real-time quantitative PCR detection of transgenes. *J. Agric. Food Chem.* **2004**, *52*, 3372–3377.

(15) Hernandez, M.; Esteve, T.; Pla, M. Real-time polymerase chain reaction based assays for quantitative detection of barley, rice, sunflower, and wheat. *J. Agric. Food Chem.* **2005**, *53*, 7003–7009.

(16) Yang, L.; Pan, A.; Jia, J.; Ding, J.; Chen, J.; Huang, C.; Zhang, C.; Zhang, D. Validation of a tomato specific gene, LAT52, used as an endogenous reference gene in qualitative and real-time quantitative PCR detection of transgenic tomatoes. *J. Agric. Food Chem.* **2005**, *53*, 183–190.

(17) Hernandez, M.; Duplan, M. N.; Berthier, G.; Vaitilingom, M.; Hauser, W.; Freyer, R.; Pla, M; Bertheau, Y. Development and comparison of four real-time polymerase chain reaction systems for specific detection and quantification of *Zea mays*. *J. Agric. Food Chem.* **2004**, *52*, 4632–4637.

(18) Mayu, I.; Satomi, Y.; Hirohito, Y.; Katsuyuki, H.; Hideo, K.; Takashi, K.; Satoshi, F.; Hiroshi, A.; Tamio, M.; Akhihiro, H. Development of taxon-specific sequences of common wheat for the detection of genetically modified wheat. *J. Agric. Food Chem.* **2005**, *53*, 6294–6300.

(19) When-Tao, X.; Kun-Lun, H.; Ying, W.; Hong-Xing, Z.; Yun-Bo, L. A cotton-specific gene, stearoyl-ACP desaturase, used as a reference gene for qualitative and quantitative polymerase chain reaction detection of genetically modified organisms. *J. Sci. Food Agric.* **2006**, *86*, 1103–1109.

(20) Yang, L.; Jianxu, C.; Huang, C.; Liu, Y.; Jia, S; Pan, L.; Zhang, D. Validation of a cotton - specific gene, sad1, used as an endogenous reference gene in qualitative and real-time quantitative PCR detection of transgenic cottons. *Plant Cell Rep.* **2005**, *24*, 237–245.

(21) *ISO/DIS/21571, Foodstuffs—Method of Analysis for the Detection of Genetically Modified Organisms and Derived Products—Nucleic Acid Extraction*; International Organization for Standardization: Geneva, Switzerland, 2002.

(22) Délye, C.; Matéjicek, A.; Gasquez, J. PCR-based detection of resistance to acetyl-CoA carboxylase-inhibing herbicides in blackgrass (*Alopecurus myosuroides* Huds) and ryegrass (*Lolium rigidum* Gaud). *Pest Manage. Sci.* **2002**, *58*, 474–478.

(23) Taberlet, P.; Gielly, L.; Pautou, G.; Bouvet, J. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.* **1991**, *17*, 1105–1109.

(24) International Rice Genome Sequencing Project. The map-based of the rice genome. *Nature* **2005**, *436*, 793–800.

(25) Liberles, D. A.; Schreiber, D. R.; Govindarajan, S.; Chamberlin, S. G.; Benner, S. A. The Adaptive Eevolution Database (TAED). *Gen. Biol.* **2001**, *8*, 281–286.

(26) Roth, C.; Betts, M. J.; Steffansson, P.; Saelensminde, G.; Liberles, D. A. The Adaptive Evolution Database (TAED): A phylogeny based tool for comparative genomics. *Nucleic Acids Res.* **2005**, *33*, 495–497.

(27) Nielsen, T. H.; Rung, J. H.; Villadsen, D. Fructose-2-6-biphosphate: A traffic signal in plant metabolism. *TRENDS Plant Sci.* **2004**, *9*, 556–563.

(28) International Organization for Standardization. *International Standard (ISO) 5725. Accuracy (Trueness and Precision) of Measurement Methods and Results*; International Organization for Standardization: Geneva, Swizerland, 1994.

(29) International Organization for Standardization. *ISO/DIS 24276 Foodstuffs—Methods of Analysis for the Detection of Genetically Modified Organisms and Derived Products—General Requirements and Definitions ISO*; International Organization for Standardization: Geneva, Switzerland, 2006.

(30) Arumuganathan, K.; Earle, E. D. Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* **1991**, *9*, 208–218.

(31) Litao, Y.; Jiayu, D.; Chngmei, Z.; Junwei, J.; Haibo, W.; Wenxuan, L.; Dabing, Z. Estimating the copy number of transgenes in transformed rice by real-time quantitative PCR. *Plant Cell Rep.* **2005**, *23*, 759–763.

(32) P. Song, C. Q; Cai, M.; Skokut, M.; Kosegi, B. D.; Petolino, J. F. Quantitative real-time PCR as a screening tool for estimating transgene copy number in WHISKERS—Derived transgenic maize. *Plant Cell Rep.* **2002**, *20*, 948–954.

(33) Anwaar, A.; Shahina, B. M.; Syed, A. H.; Mariam, B. S. Determination of cryIAb and cryIAc copy number in transgenic Basmati 370 rice (*Oryza sativa* L.) plants using real-time PCR and its comparaison with southern blot. *J. Biol. Sci.* **2005**, *5*, 283–288.

(34) Burner, B.; Klaus, G.; Baldwin, I. Two-fold differences are the detection limit for determining transgene copy numbers in plants by real-time PCR. *BMC Biotechnol.* **2004**, 4–14.

(35) Knapp, S. Tobacco to tomatoes: A phylogenetic perspective on fruit diversity in the Solanaceae. *J. Exp. Bot.* **2002**, *53*, 2001–2022.

(36) Fridman, E.; Carrari, F.; Liu, Y. S.; Fernie, A. R.; Zamir, D. Zooming in on a quantitative trait for tomato yield using interspecific introgression. *Science* **2004**, *305*, 1786–1789.

(37) Barocka, K. H. Z. Und Futterruben (*Beta vulgaris*). *Lehrbuch der Züchtung Landwirtschaftlicher Kulturpflazen*;: 1985; pp 245–287.

(38) Van Geyt, J. P.C.; Lange, W.; Oleo, M.; Debock, T. S. M. Natural variation within the genus *Beta* and its possible use for breeding sugar-beet—A review. *Euphytica* **1990**, *49*, 57–76.

(39) U, N. Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Jpn. J. Bot.* **1935**, *7*, 389–452.

(40) Schmidt, A. M.; Rott, M. E. Real-time polymerase chain reaction (PCR) quantitative detection of *Brassica napus* using a locked nucleic acid taqman probe. *J. Agric. Food Chem.* **2006**, *54*, 1158–1165.

Received for review May 15, 2007. Revised manuscript received July 27, 2007. Accepted July 31, 2007. Financial support from the DGCCRF (Direction Générale de la Consommation, de la Concurrence et de la Répression des Fraudes) (Contract A02543, GMO convention no. 1009) and by the European Commission through the Integrated Project "Co-Extra" (Contract 007158, under the 6th Framework Programme, priority 5, food quality and safety) is gratefully acknowledged.

JF071429D