Eur J Plant Pathol (2011) 129:89–102
DOI 10.1007/s10658-010-9695-9

Chorismate mutase: an alternatively spliced parasitism gene and a diagnostic marker for three important *Globodera* nematode species

Hang Yu · Demosthenis Chronis · Shunwen Lu · Xiaohong Wang

Accepted: 27 September 2010 / Published online: 22 October 2010

Abstract The chorismate mutase gene is widely distributed in both cyst and root-knot nematode species and believed to play a critical role in nematode parasitism. In this study, we cloned a new chorismate mutase gene (*Gt-cm-1*) from *Globodera tabacum* and further characterized the gene structure in both *G. tabacum* and *G. pallida*, a closely related species of *G. rostochiensis*. The genomic clones of chorismate mutase genes from these two species were found to contain three introns with the second intron having unusual 5' and 3' splice sites. A previous study revealed that the chorismate mutase gene from *G. rostochiensis* is subject to alternative splicing through retention of intron 2, a process that allows for the generation of multiple mRNA transcripts from a single gene. As expected, we discovered that alternative splicing of the chorismate mutase gene is a conserved event in three *Globodera* species, supporting an important role of alternative splicing in regulating chorismate mutase gene function in plant parasitism by these nematodes. In addition to the potential suboptimal 5' and 3' splice sites and the small size of intron 2, detailed sequence analysis also identified candidate *cis*-acting elements that might be responsible for regulating intron retention of *Globodera* chorismate mutase genes. Based on genomic sequence variations observed, we developed TaqMan qPCR assays that provided a highly specific and sensitive identification of each *Globodera* species, revealing a new application of using the chorismate mutase gene as a valuable diagnostic marker for plant-parasitic nematodes.

Keywords Alternative splicing · intron retention · TaqMan qPCR

Abbreviations

AS	alternative splicing
CM	chorismate mutase
ESE	exonic splicing enhancer
Gr-cm-1	*Globodera rostochiensis* chorismate mutase gene
Gp-cm-1	*Globodera pallida* chorismate mutase gene
Gt-cm-1	*Globodera tabacum* chorismate mutase gene
LNA	locked nucleic acids
PCN	potato cyst nematode

H. Yu · X. Wang (✉)
Department of Plant Pathology and Plant-Microbe Biology,
Cornell University,
Ithaca, NY 14853, USA
e-mail: xiaohong.wang@ars.usda.gov

D. Chronis · X. Wang
Robert W. Holley Center for Agriculture and Health,
USDA-ARS,
Ithaca, NY 14853, USA

S. Lu
Northern Crop Science Laboratory,
USDA-ARS,
1307 18th Street North,
Fargo, ND 58105, USA

qPCR quantitative real-time polymerase chain reaction
TCN tobacco cyst nematode

Introduction

Homologues of the chorismate mutase (*CM*) gene have been found to be widely distributed in both cyst and root-knot nematodes including the two species of potato cyst nematode (PCN) *Globodera rostochiensis* and *G. pallida* (Jones et al. 2003; Lu et al. 2008). Due to its exclusive expression within nematode's esophageal gland cells and its possible function in altering host cellular metabolism, chorismate mutase secreted from nematodes is believed to play a critical role in nematode parasitism (Davis et al. 2004; Doyle and Lambert 2003). A recent study of the *CM* gene (*Gr-cm-1*) from *G. rostochiensis* revealed that *Gr-cm-1* is subject to alternative splicing (AS) through retention of intron 2 leading to two different splicing transcripts (*Gr-cm-1* and *Gr-cm-1-IRII*) that encode a functional CM and a truncated non-functional CM (Lu et al. 2008). Alternative pre-mRNA splicing is an important mechanism for increasing eukaryotic transcriptome and proteome diversity by producing multiple mRNA transcripts from a single gene (Blencowe 2006). Although AS has been extensively studied in many eukaryotic organisms including mammals (Blencowe 2006; Soller 2006) and the free-living nematode *Caenorhabditis elegans* (Zahler 2005), until today only one example of AS has been reported in plant-parasitic nematodes (Lu et al. 2008).

The two PCN species are significant quarantine pests causing serious problems in potato production and trade worldwide. In recent years new detections of PCN have been reported in both the U.S. (Skantar et al. 2007) and Canada (Sun et al. 2007), revealing an increasing threat of these pests to the potato industry in North America. The tobacco cyst nematode (TCN) *G. tabacum*, a closely related species of PCN, is also an important nematode pest which has been found in tobacco production areas of the U.S., Canada, Mexico, and many European countries (Bélair and Miller 2006; Marché et al. 2001; Syracuse et al. 2004).

Traditional diagnosis of *Globodera* species based on morphology and morphometric characteristics requires a high degree of skill and is time-consuming. Molecular diagnostic methods utilizing ribosomal DNA (rDNA) sequences have been developed for identifying these three important *Globodera* species (Bates et al. 2002; Bulman and Marshall 1997; Madani et al. 2008; Skantar et al. 2007; Thiéry and Mugniéry 1996). The species-specific PCR primers developed by Bulman and Marshall (1997) have been found to be useful in differentiating PCN species and species-specific PCR primers for TCN identification were also recently made available (Skantar et al. 2007). More recently, quantitative real-time PCR (qPCR) assays that include melting curve analysis of PCR products or TaqMan technology utilizing locked nucleic acids (LNA)-modified primers and probes, have been developed for identifying *Globodera* species (Bates et al. 2002; Madani et al. 2008). qPCR-based diagnostics offer rapid, specific, and sensitive detection compared to conventional PCR assays that require post-PCR gel analysis. However, the potential of qPCR-based strategies by targeting genes other than rDNA sequences for identifying *Globodera* species has yet to be explored.

Because of the involvement of CM in nematode parasitism and the unique regulation of the *CM* gene discovered in *G. rostochiensis* (Lu et al. 2008), we cloned a *CM* homologue from *G. tabacum* and further characterized the *CM* gene structure in both *G. tabacum* and *G. pallida*. Similar to *G. rostochiensis*, we found evidence for alternative splicing of *CM* genes through intron retention in these two *Globodera* species. A close examination of *Globodera CM* genes identified putative *cis*-acting elements that might be responsible in regulating intron retention. Alignment of different genomic sequences of *Globodera CM* genes identified regions unique to each of the three *Globodera* species which we exploited to develop highly sensitive and reliable TaqMan qPCR assays for rapid species-specific differentiation. To our knowledge, this is the first report of using an esophageal gland expressed parasitism gene as a genetic marker for diagnosis of plant-parasitic nematodes.

Materials and methods

Nematode populations and genomic DNA extraction

Six populations of potato cyst nematode (PCN) *Globodera rostochiensis*, seven populations of potato

cyst nematode *G. pallida*, two populations of tobacco cyst nematode (TCN) *G. tabacum*, and one population of *Heterodera glycines* and *H. schachtii* (Table 1) were used in this study. Genomic DNA from nematode populations of Gr-Ama, Gr-NF, Gr-Ro1-NL, Gr-Ro5, Gp-NF, Gp-Pa2, Gp-297, Gp-286, Gp-Pa2/3, and Gp-Pa1 was obtained from collaborators. Hatched second-stage juveniles of Gr-Ro1, Gr-Ro2, *H. glycines*, and *H. schachtii* were used for DNA extraction according to Lu et al. (2008) and cysts of Gp-ID were used for DNA extraction as follows: one to ten dried cysts were crushed and homogenized in 150 μl of lysis buffer (200 mM Tris-HCl, pH 8.5, 250 mM NaCl, 25 mM EDTA, 0.5% SDS) by vortexing for 2 min. Half volume of 3 M sodium acetate (pH 5.2) was added to the homogenate, then incubated at -20°C for 10 min. Supernatant was collected by centrifugation at 13,200 rpm for 5 min, then incubated with 1 volume of isopropanol on ice for 30 min. Genomic DNA was precipitated by centrifugation at 13,200 rpm for 15 min and the pellet was washed twice with 70% ice-cold ethanol and once with 100% ice-cold ethanol. The genomic DNA was resuspended in 15 to 20 μl of TE buffer.

Gene cloning

Parasitic stages of *G. tabacum* extracted from infected tobacco roots (Wang et al. 2001) and infected potato root segments that contained *G. pallida* at different developmental stages were used for mRNA extraction as described previously (Lu et al. 2008). First-strand cDNA synthesis was conducted using the 3'RACE system (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. Full-length chorismate mutase (*CM*) genes were amplified by PCR using primers GrCM1-atgF and GrCM1-tgaR (Table 2) corresponding to the 5'- and 3'-end of the open reading frame (ORF) of the *Gr-cm-1* gene

Table 1 Nematode populations used in this study

Species	Population abbreviation	Origin	No. of independent samples	*CM* sequence accession no.
Globodera rostochiensis				
	Gr-Ro1	New York, USA	2	EF437152
				EF437153
	Gr-Ro2	New York, USA	2	
	Gr-Ama	Quebec, Canada	3	
	Gr-NF	Newfoundland, Canada	1	
	Gr-Ro1-NL	The Netherlands	1	
	Gr-Ro5	The Netherlands	1	
G. pallida				
	Gp-ID	Idaho, USA	3	HM148923
				HM148924
				HM148925
	Gp-NF	Newfoundland, Canada	1	
	Gp-Pa2	The Netherlands	1	
	Gp-297	South Peru	1	
	Gp-286	South Peru	1	
	Gp-Pa2/3	Chavornay, Switzerland	1	
	Gp-Pa1	Duddingston, UK	1	
G. tabacum				
	Gt-MO	unknown	1	HM148926
				HM148927
	Gt-NC	North Carolina, USA	1	
Heterodera glycines	Hg	North Carolina, USA	1	
H. schachtii	Hs	New York, USA	1	

Table 2 Primers and probes used in this study

Name	Sequence (5' to 3')
Primers	
GrCM1-atgF	ATGAATTTGTTGGTCGTTCCGT
GrCM1-tgaR	TCATTCATTCAGCAGTTTCTTG
GrpCM1-167F	CAAATAATAGGCCAAATTGGAT
GrCM1ab-408R	CTTCAGTCCAAGGCTAATTCTC
GpCM1-1551F	TGAAGCTTTCGGCAGTTAT
GpCM1-1811R	GGTGACCGTCTGCAAGT
GtCM1-915F	GTCATCGGAGGAAGTCATT
GtCM1-1153R	CAAAAAATGACCAAATCGAA
Probes	
GrCM1-308P	FAM-CATTCGACGCTTGCCTTTCGC-TAMRA
GpCM1-1692P	FAM-TCCCTAACGAACTGAGGCTTACCG-TAMRA
GtCM1-1010P	FAM-AACGGAAACGACAACGGGAAA-TAMRA

(GenBank accession number EF437154). GrCM1-atgF and GrCM1-tgaR were also used to amplify the corresponding genomic sequences of the *CM* gene from *G. pallida* or *G. tabacum* genomic DNA. PCR cycling conditions were 95°C for 5 min, followed by 35 cycles of 95°C for 40 s, 50°C for 40 s, 72°C for 3 min, and a final reaction of 72°C for 10 min. Amplified PCR products were cloned into the pGEM-T vector (Promega, Madison, WI, USA) and sequenced at the Cornell University DNA Sequencing Facility.

Southern blot analysis

Approximately 10 μg of TCN genomic DNA was digested with *Bam*HI, *Eco*RI, or *Hin*dIII, separated by electrophoresis on a 0.7% agarose gel, transferred onto a positively charged nylon membrane (Roche Applied Science, Indianapolis, IN, USA), fixed using a Spectrolinker XL-1,000 UV crosslinker (Spectronics Corporation, Westbury, NY, USA), and hybridized following a standard protocol (Sambrook et al. 1989). A digoxigenin (DIG)-labelled *Gt-cm-1* probe corresponding to the ORF of *Gt-cm-1* was generated by PCR with GrCM1-atgF and GrCM1-tgaR primers using the PCR DIG Probe Synthesis Kit (Roche Applied Science, Indianapolis, IN, USA). About 10 ng of DIG-labelled probe per ml was used for hybridization. Hybridization was performed at 65°C for 14 h and the membrane was further processed to reveal hybridization signals as previously described (Lu et al. 2008).

Specificity of TaqMan quantitative real-time PCR (qPCR) assays

Alignment of *CM* genes from *G. rostochiensis*, *G. pallida*, and *G. tabacum* identified regions unique to each of the three species and specific TaqMan probes targeting these unique regions were developed. A TaqMan qPCR assay was used to identify each *Globodera* species. Primers GrpCM1-167F and GrCM1ab-408R and probe GrCM1-308P (Table 2) were used for the specific detection of *Gr-cm-1* from *G. rostochiensis*. Primers GpCM1-1551F and GpCM1-1811R and probe GpCM1-1692P (Table 2) were used for the specific detection of *Gp-cm-1* from *G. pallida*. Primers GtCM1-915F and GtCM1-1153R and probe GtCM1-1010P (Table 2) were used for the specific detection of *Gt-cm-1* from *G. tabacum*. The specificity of each TaqMan qPCR assay was tested against DNA samples extracted from six different populations of *G. rostochiensis* (Table 1), seven different populations of *G. pallida* (Table 1), two different populations of *G. tabacum* (Table 1), one population of *Heterodera glycines* and *H. schachtii* (Table 1), and an artificially-mixed population of three *Globodera* species. qPCR was carried out in an iCycler iQ Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA, USA) as previously described (Lu et al. 2008). Briefly, the assay was conducted in a 96-well plate in a 25 μl reaction volume containing iQ Supermix (50 mM KCl, 20 mM Tris-HCl, pH 8.4, 0.8 mM dNTPs, 0.375 U iTaq DNA polymerase, 3 mM $MgCl_2$, and stabilizers) (Bio-Rad),

500 nM of both primers, 500 nM probe, and 1 μl of genomic DNA. PCR reactions with no DNA template were also included as a negative control. qPCR was started with an iTaq DNA polymerase activation and DNA denaturation step (95°C for 3 min), then followed by 40 cycles of 95°C for 20 s and 60°C for 40 s. All qPCR assays consisted of three technical replicates and at least two independent qPCR assays were conducted for nematode populations of Gr-Ro1, Gr-Ro2, Gr-Ama, and Gp-ID (Table 1). Data were analyzed using the iCycler iQ Real-Time PCR Detection System Software version 3.0a (Bio-Rad). Gel electrophoresis analysis was also performed to verify that only one expected PCR product specific to each species was amplified. All TaqMan probes were synthesized at Sigma-Aldrich (St. Louis, MO, USA) that contained a fluorescence reporter dye (6-carboxyfluorescein or FAM) and a quencher dye (6-carboxy-tetramethyl-rhodamine or TAMRA) at their 5'- and 3'-ends, respectively.

Sensitivity of TaqMan qPCR assays

To determine the sensitivity of each TaqMan qPCR assay, genomic DNA extracted from each *Globodera* species was quantified with a spectrophotometer (Shimadzu Corporation, Kyoto, Japan) and ten-fold serial dilutions were made and used in TaqMan qPCR assays to generate standard curves according to a standard protocol (User Bulletin 2, ABI PRISM 7700 Sequence Detection System, Applied Biosystems, Foster City, CA, USA). Three replicates for each dilution were done in each qPCR assay. Two independent experiments were conducted and similar results were obtained.

Results

Chorismate mutase cDNA clones from *G. pallida* and *G. tabacum*

A recent study reported only one *Gp-cm-1* cDNA sequence (GenBank accession number AJ457834) (Jones et al. 2003). However, sequencing of several random clones from PCR reactions with first-strand cDNA generated from *G. pallida* mRNA identified a second cDNA clone with significant sequence similarity to the reported *Gp-cm-1* cDNA sequence (Jones et al. 2003). This cDNA clone, named *Gp-cm-1-IRII* (GenBank accession number HM148921), was 98.5% identical to *Gp-cm-1* in the area of sequence overlap, but contained a 93-bp insertion in the region (Fig. 1a). The *Gp-cm-1-IRII* cDNA contained an open reading frame (ORF) of 366 bp (Fig. 1a) encoding a deduced truncated protein (Gp-CM-1t) of 121 amino acids lacking the conserved CM domain (Lu et al. 2008) due to the introduction of a stop codon caused by the insertion of the 93-bp fragment (Fig. 1a).

Sequencing of random clones from PCR reactions with first-strand cDNA generated from *G. tabacum* mRNA identified two cDNA clones (named *Gt-cm-1* and *Gt-cm-1-IRII*) that revealed significant similarity to *CM* genes from *G. rostochiensis* and *G. pallida*. The *Gt-cm-1* cDNA (GenBank accession number HM148920) is 97.1% and 94.1% identical to *Gr-cm-1* (GenBank accession number EF437154) and *Gp-cm-1* in the ORF region and encodes a deduced protein (Gt-CM-1) with an N-terminal signal peptide (Nielsen et al. 1997) and a conserved CM domain (Fig. 1b). The *Gt-cm-1-IRII* cDNA (GenBank accession number HM148922) is identical to *Gt-cm-1* in the region of sequence overlap except for having an 89-bp insertion in the region. The *Gt-cm-1-IRII* cDNA encodes a deduced truncated protein (Gt-CM-1t) of 104 amino acids that lacks the conserved CM domain due to the generation of a stop codon induced by the insertion of the 89-bp fragment (Fig. 1b).

Genomic clones and sequence analysis

Three genomic sequences corresponding to *Gp-cm-1* and *Gp-cm-1-IRII* were obtained by sequencing of more than twenty random clones from PCR amplifications with genomic DNA from *G. pallida*. These three genomic clones (named *Gp-cm-1A*, *Gp-cm-1B*, and *Gp-cm-1C*) (GenBank accession numbers HM148923, HM148924, and HM148925) match to the *Gp-cm-1* cDNA except that three introns were identified as indicated in Fig. 1a. The *Gp-cm-1* genomic sequence (GenBank accession number AJ487621) previously reported by Jones et al. (2003) was found to contain only two introns. However, extensive sequence analysis of random *Gp-cm-1* genomic clones revealed that they all contain three introns. The three introns in *Gp-cm-1A*, *Gp-cm-1B*, and *Gp-cm-1C* follow the 'GU-AG' rule for *cis*-splicing. However, similar to what was discovered in *Gr-cm-1A* and *Gr-cm-1B* (Lu et al.

2008), the 5' and 3' splice sites of intron 2 were found to be divergent differing in three nucleotides compared to the respective consensus sequences (AG/GURAGUUU and UUUUCAG/R) identified in *C. elegans* (Blumenthal and Steward 1997) (Table 3). There are only three single nucleotide polymorphisms (SNPs) in the second intron (intron 2) of the three *Gp-cm-1* genomic sequences, however, obvious sequence variations were found in the first (intron 1) and the third intron (intron 3) of these genomic sequences. In addition to many SNPs in intron 1, both *Gp-cm-1B* and *Gp-cm-1C* have an insertion of approximately 50 bp in the region whereas *Gp-cm-1A* lacks this insertion (Fig. 1a). Furthermore, compared to *Gp-cm-1A* and *Gp-cm-1B*, *Gp-cm-1C* contains a stretch of very dissimilar sequence in intron 3 (Fig. 1a). These genomic sequences may represent different alleles of *Gp-cm-1* or different gene copies. Sequence analysis further discovered that the 93-bp insertion found in the *Gp-cm-1-IRII* cDNA was derived from the entire second intron of the genomic sequence, revealing that *Gp-cm-1-IRII* is a splice variant generated from alternative splicing of the *Gp-cm-1* pre-mRNA.

Two genomic sequences (named *Gt-cm-1A* and *Gt-cm-1B*) (GenBank accession numbers HM148926 and HM 148927) corresponding to *Gt-cm-1* and *Gt-cm-1-IRII* were obtained by sequencing of about ten random clones from PCR amplifications with genomic DNA from *G. tabacum*. These genomic clones match to the *Gt-cm-1* cDNA except that three introns were identified as indicated in Fig. 1b. The three introns in *Gt-cm-1A* and *Gt-cm-1B* follow the 'GU-AG' rule for *cis*-splicing. Similarly, the 5' and 3' splice sites of intron 2 were divergent differing in either three or five nucleotides compared to the respective consensus sequences identified in *C. elegans* (Table 3). *Gt-cm-1A* and *Gt-cm-1B* are highly similar with each other and only a few nucleotide variations were identified in intron 1 and intron 3. Further sequence analysis revealed that the 89-bp insertion found in the *Gt-cm-1-IRII* cDNA was derived from the entire second intron of the genomic sequence, revealing that *Gt-cm-1-IRII* is also a splice variant generated from alternative splicing of the *Gt-cm-1* pre-mRNA.

Southern blot analysis was used to investigate the gene copy number of *Gt-cm-1* in *G. tabacum*. *Gt-cm-1* genomic clones did not contain a *Bam*HI restriction site, but did contain one *Eco*RI, and two *Hin*dIII digestion sites (Fig. 1b). Although two genomic sequences of *Gt-cm-1* were identified, when DIG-labeled *Gt-cm-1* cDNA probe was hybridized to a membrane containing *Bam*HI, *Eco*RI, or *Hin*dIII-digested genomic DNA, only one fragment was detected on the membrane with *Bam*HI digestion, and two and three fragments were detected with *Eco*RI or *Hin*dIII digestion (Fig. 2), indicating that *Gt-cm-1* exists as a single-copy gene in the *G. tabacum* genome. Thus, the two *Gt-cm-1* genomic sequences may represent different alleles of *Gt-cm-1*.

Fig. 1 Schematic diagrams of the genomic organization of *CM* genes from *Globodera pallida* **a**, *G. tabacum* **b**, and *G. rostochiensis* **c** (Lu et al. 2008) and their mature mRNA transcripts. Exons are indicated by boxes; stippled boxes indicate the N-terminal signal peptide and black boxes indicate the chorismate mutase domain. Introns are indicated by solid lines; stippled lines indicate sequence variations among *Gp-cm-1* sequences. Caret lines indicate exon-exon junctions in mRNA transcripts. Numbers in the boxes and below the lines indicate the sizes (in base pairs) of the exons and introns, respectively. Numbers between arrow lines indicate the sizes of the open reading frames of *Gp-cm-1*, *Gp-cm-1-IRII*, *Gt-cm-1*, and *Gt-cm-1-IRII*, respectively. Double arrows indicate the regions targeted for the development of species-specific TaqMan probes. In *Gp-cm-1-IRII* and *Gt-cm-1-IRII*, the alternative exon 2 also includes partial sequence of intron 2 (35 bp for *Gp-cm-1-IRII* and 11 bp for *Gt-cm-1-IRII*). The rest region of each transcript corresponding to exon 3 and exon 4 becomes noncoding exons (shaded) because of the presence of a stop codon (indicated by asterisks) introduced by retention of intron 2. The one *Eco*RI and two *Hin*dIII digestion sites are also marked on *Gt-cm-1A* and *Gt-cm-1B*

Unique sequences identified in *Globodera CM* genes

A close examination of *Globodera CM* sequences revealed sequence variations predominantly in intron 1 and intron 3. A short sequence stretch that can be used for distinguishing *G. rostochiensis CM* alleles from other *CM* sequences was identified in intron 1 (Fig. 1c), which was found to be conserved within *G. rostochiensis* populations but divergent from the other two *Globodera* species. Similarly, a short sequence stretch that can be used for distinguishing *G. pallida CM* sequences from other *CM* alleles was identified in intron 3 (Fig. 1a). Further sequence comparison also discovered a short sequence stretch in exon 2 that can be used for distinguishing *G. tabacum CM* alleles from other *CM* sequences (Fig. 1b).

Although the information present in canonical splice signals (e.g., 5' and 3' splice sites) is generally essential for splicing, auxiliary splicing elements that are typically located near the splice-junction are also

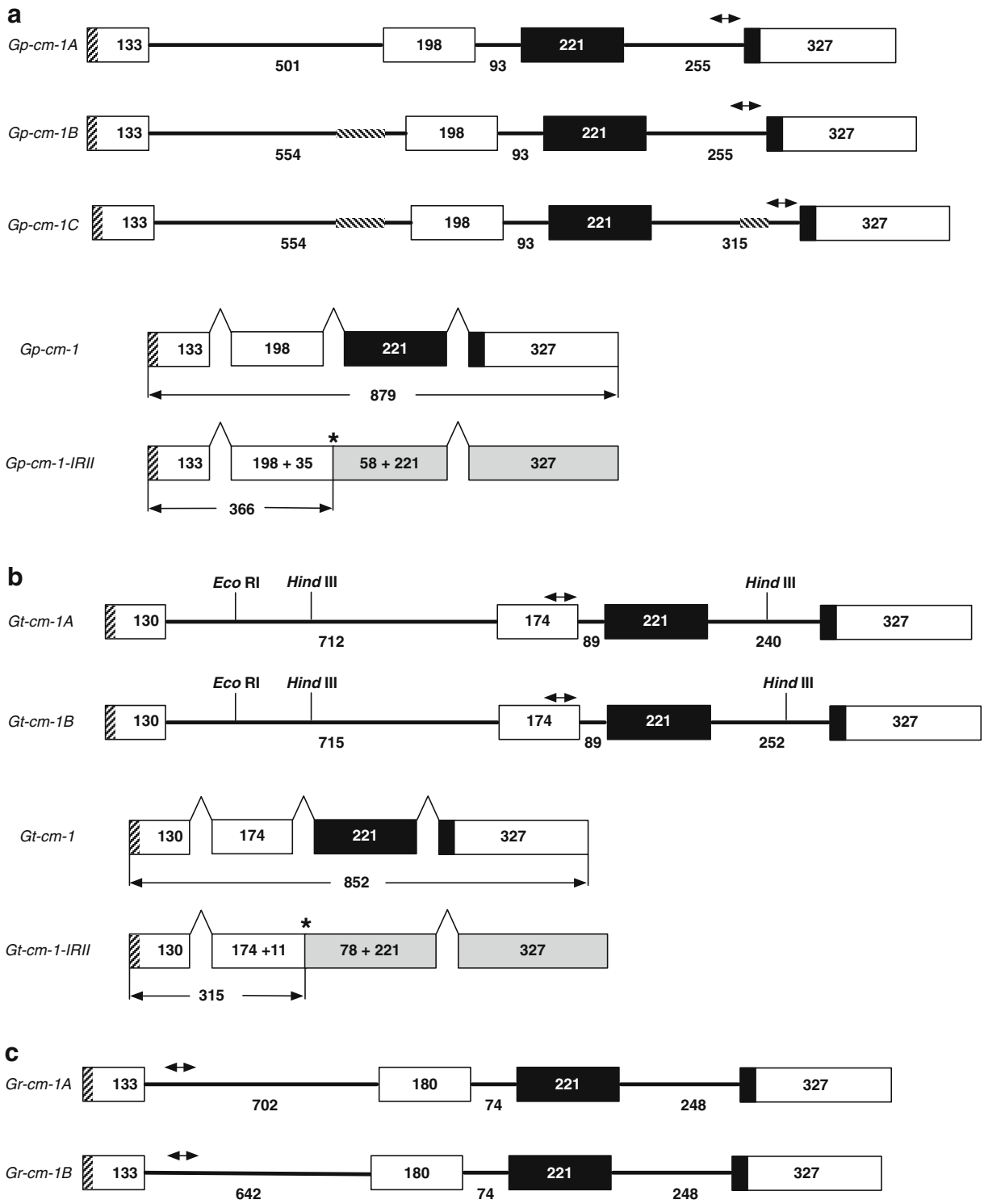
a
Gp-cm-1A
133
198
221
327
501
93
255
Gp-cm-1B
133
198
221
327
554
93
255
Gp-cm-1C
133
198
221
327
554
93
315
Gp-cm-1
133
198
221
327
879
Gp-cm-1-IRII
133
198 + 35
58 + 221
327
366
b
Eco RI
Hind III
Hind III
Gt-cm-1A
130
174
221
327
712
89
240
Eco RI
Hind III
Hind III
Gt-cm-1B
130
174
221
327
715
89
252
Gt-cm-1
130
174
221
327
852
Gt-cm-1-IRII
130
174 +11
78 + 221
327
315
c
Gr-cm-1A
133
180
221
327
702
74
248
Gr-cm-1B
133
180
221
327
642
74
248

Table 3 Splice site sequences of *C. elegans* and intron 2 of *Globodera CM* genes

Species	5' splice site	3' splice site
C. elegans	AG/GURAGUUU	UUUUCAG/R
G. rostochiensis cm-1A/B	AG/GUA*C*G*GUC*	U*GAC*CAG/A
G. pallida cm-1A/B/C	AG/GUA*C*G*GUC*	U*CAC*CAG/A
G. tabacum cm-1A/B	AG/GUA*CUCAC*	U*GAC*CAG/A

Nucleotides different from those in the consensus sequences of *C. elegans* are indicated in italic

R represents A or G

important for ensuring accurate and efficient splicing (Cartegni et al. 2003; Voelker and Berglund 2010). A close examination of sequences adjacent to the splice-junctions of intron 2 identified several purine-rich elements resembling splicing enhancers previously identified in animal genes (Dye et al. 1998; Lejeune et al. 2001; Mercado et al. 2005) (Fig. 3). ESEfinder (http://exon.cshl.edu/ESE/) is a web-based resource for predicting putative exonic splicing enhancers (ESE) in human genes (Cartegni et al. 2003). Since the mechanism of alternative splicing regulation is suggested to be well conserved in metazoans (Kabat et al. 2006), we used ESEfinder to identify putative ESEs in sequences around intron 2 splice-junctions. Three high-scoring putative ESEs located in exon 3 downstream of the 3' splice site of intron 2 were identified and found to be conserved among *Globodera CM* genes (Fig. 3).

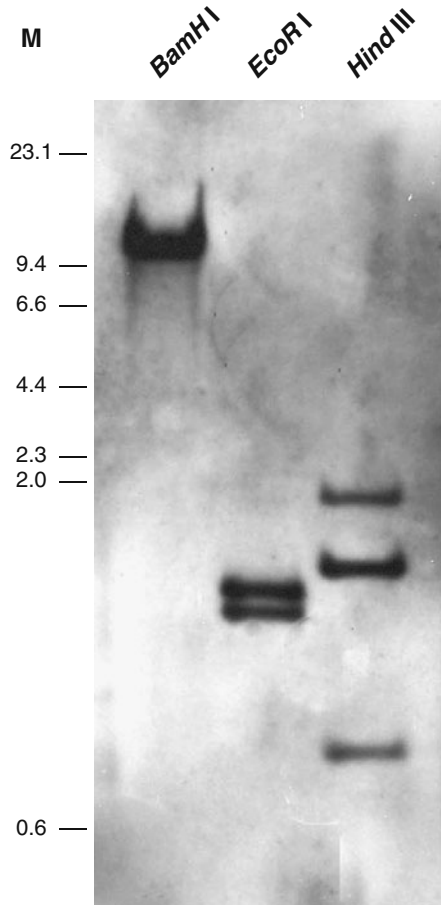

Fig. 2 Southern blot analysis of the *Gt-cm-1* gene. Genomic DNA from *Globodera tabacum* was digested with *Bam*HI, *Eco*RI, or *Hin*dIII and hybridized with the *Gt-cm-1* cDNA probe. Lane M: DIG-labeled molecular weight marker in kilobases

Specific detection of *Globodera* species

We developed species-specific TaqMan qPCR assays for identifying each *Globodera* species by targeting the short sequence stretches that were found to be unique to each *Globodera CM* gene. The specificity of each TaqMan qPCR assay was confirmed by testing six different populations of *G. rostochiensis*, seven different populations of *G. pallida*, and two different populations of *G. tabacum* originating from diverse geographical regions (Table 1). As expected, fluorescence signals were detected for all the *G. rostochiensis* populations only when the *G. rostochiensis*-specific primers-probe set was used in the assay (Fig. 4a). Likewise, fluorescence signals were detected for all the *G. pallida* and *G. tabacum*

Fig. 3 Putative splicing enhancers identified in the sequence ▶ covering the region from exon 2 to exon 3 of *Globodera CM* genes. Exon and intron sequences are indicated in upper and lower cases, respectively. Purine-rich elements are shaded and exonic splicing enhancers predicted by ESEfinder are boxed. The amino acid sequences derived from the corresponding alternative splicing transcripts are shown under nucleotide sequences and the stop codon (marked as *bold letters* in nucleotide sequences) introduced by retention of intron 2 is indicated as an asterisk

```
Gr-cm-1      ATTACCTAACCGAAAACAACAAATGCAAGTCATCGGAGGAAGTCATTTTGCGCAAGTCCG 836
Gp-cm-1      ATTACCTAGCCGAAAACAACAAATGCAAGTCATCGGAGGAAGTCATTTTGCGCAAGTCCG 631
Gt-cm-1      ATTACCTAACCGAAAACAACAAATGCAAGTCATCGGAGGAAGTCATTTTGCGTAAGTCCG 895
Gr-CM-1t       Y  L  T  E  N  N  K  C  K  S  S  E  E  I  I  L  R  K  S  D
Gp-CM-1t       Y  L  A  E  N  N  K  C  K  S  S  E  E  I  I  L  R  K  S  D
Gt-CM-1t       Y  L  T  E  N  N  K  C  K  S  S  E  E  V  I  L  R  K  S  D

Gr-cm-1      ACTGTGCCTTCATGAAGAGCATTGAGGATGGATTCAAATTTGTTGTTGGGATGGAAGGCC 896
Gp-cm-1      ACTGTGCCTTCATGAAGAACATTGAGAATGGACTCAAATTTGTTGTTGGGATGGAAGGCC 691
Gt-cm-1      ACTGTGCCTTCATGAAGAACATTGAGGATGGATTCAAATTTGTTGTTGGGATGGAAGGCC 955
Gr-CM-1t       C  A  F  M  K  S  I  E  D  G  F  K  F  V  V  G  M  E  G  Q
Gp-CM-1t       C  A  F  M  K  N  I  E  N  G  L  K  F  V  V  G  M  E  G  Q
Gt-CM-1t       C  A  F  M  K  N  I  E  D  G  F  K  F  V  V  G  M  E  G  Q

Gr-cm-1      AAACGGAAACGGAATCGACAACGGGAAA------------------TAATATTTTTATGT 938
Gp-cm-1      AAACGGAAACGGAATCCCCAACGGAAAACGCCCCAACAGCCAACAATAATAATTTTATGT 751
Gt-cm-1      AAACGGAAACG--A----CAACGGGAAA------------------TAATATTTTTATGT 991
Gr-CM-1t       T  E  T  E  S  T  T  G  N                    N  I  F  M  C
Gp-CM-1t       T  E  T  E  S  P  T  E  N  A  P  T  A  N  N  N  N  F  M  C
Gt-CM-1t       T  E  T    T     T  G  N                    N  I  F  M  C

Gr-cm-1      GTTGTAAGCCGAATCAAGgtacggtcactaaatggaaaga-----a----a--------- 980
Gp-cm-1      GTTGTAAGCCAAATCAAGgtacggtcactaaacggaaagacgacaatttgtgaaatttat 811
Gt-cm-1      GTTGTAAGCCGAATCAAGgtac--tcactaaatggaaagacgacaatttgtgaaatttac 1049
Gr-CM-1t       C  K  P  N  Q  G  T  V  T  K  W  K  E           K
Gp-CM-1t       C  K  P  N  Q  G  T  V  T  K  R  K  D  D  N  L  *
Gt-CM-1t       C  K  P  N  Q  G  T     H  *

Gr-cm-1      aatcttttcgatttgaccatttttgc-ccgtctgtaaccattttgaccagAGACGGCCA 1039
Gp-cm-1      aatcttttcgatttggtcatttttttgggccgtctgtaaccattttcaccagAGACGGCCA 871
Gt-cm-1      aatcttttcgatttggtcatttttttg--ccctctgtaaccattttgaccagAGACGGTCA 1107
Gr-CM-1t       S  F  R  F  D  H  F  L  P   V  C  N  H  F  D  Q  R  R  P

Gr-cm-1      CTCTCTTCATTGTCGGCATGGCCAACAAACGGCTGATGTTGGCCAAAGATGTGGTTCTCT 1099
Gp-cm-1      CACTCTTTATTGTCGGCGTGGCCAACAAACGGCTGATGTTGGCCAAAGACGTGGCTCTCT 931
Gt-cm-1      CGCTCTTCATTGTCGGCGTGGCCAACAAACGGCTGATGTTGGCCAAAGACGTGGTTCTCT 1167
Gr-CM-1t     L  S  S  L  S  A  W  P  T  N  G  *

Gr-cm-1      ACAAGTACATCAACAACAATAGCATTGACGATTTCGAGCGGGAAAAGGTTGTGCTGCAAA 1159
Gp-cm-1      ACAAGTACATCAACAACAATAGCATTGACGATTTTGAGCGTGAAAAGGTTGTGTTGCAAA 991
Gt-cm-1      ACAAGTACATCAACAACAATAGCATTGACGATTTTGAACGGGAAAAGGTTGTGCTGCAAA 1227

Gr-cm-1      ATGTTTTGGCTCAGGCGAAGAGTGCCGGGATAAGCGACAACTACGGGGAGCCGTTTTTCC 1219
Gp-cm-1      ATGTTTTGGCTCAAGCGAACAGTGCCGGCATAAGCGACAACTACGGGGAGCCGTTCTTCC 1051
Gt-cm-1      ATGTTTTGGCTCAGGCGAACAGTGCCGGCATAAGCGACAACTACGGGGAGCCGTTCTTCC 1287

Gr-cm-1      AAGACCAAATGGACGCTAACAAAGTCATTCAG 1251
Gp-cm-1      AAGACCAAATGGACGCTAACAAAGTCATTCAG 1083
Gt-cm-1      AAGACCAAATGGACGCTAACAAAGTCATTCAG 1319
```

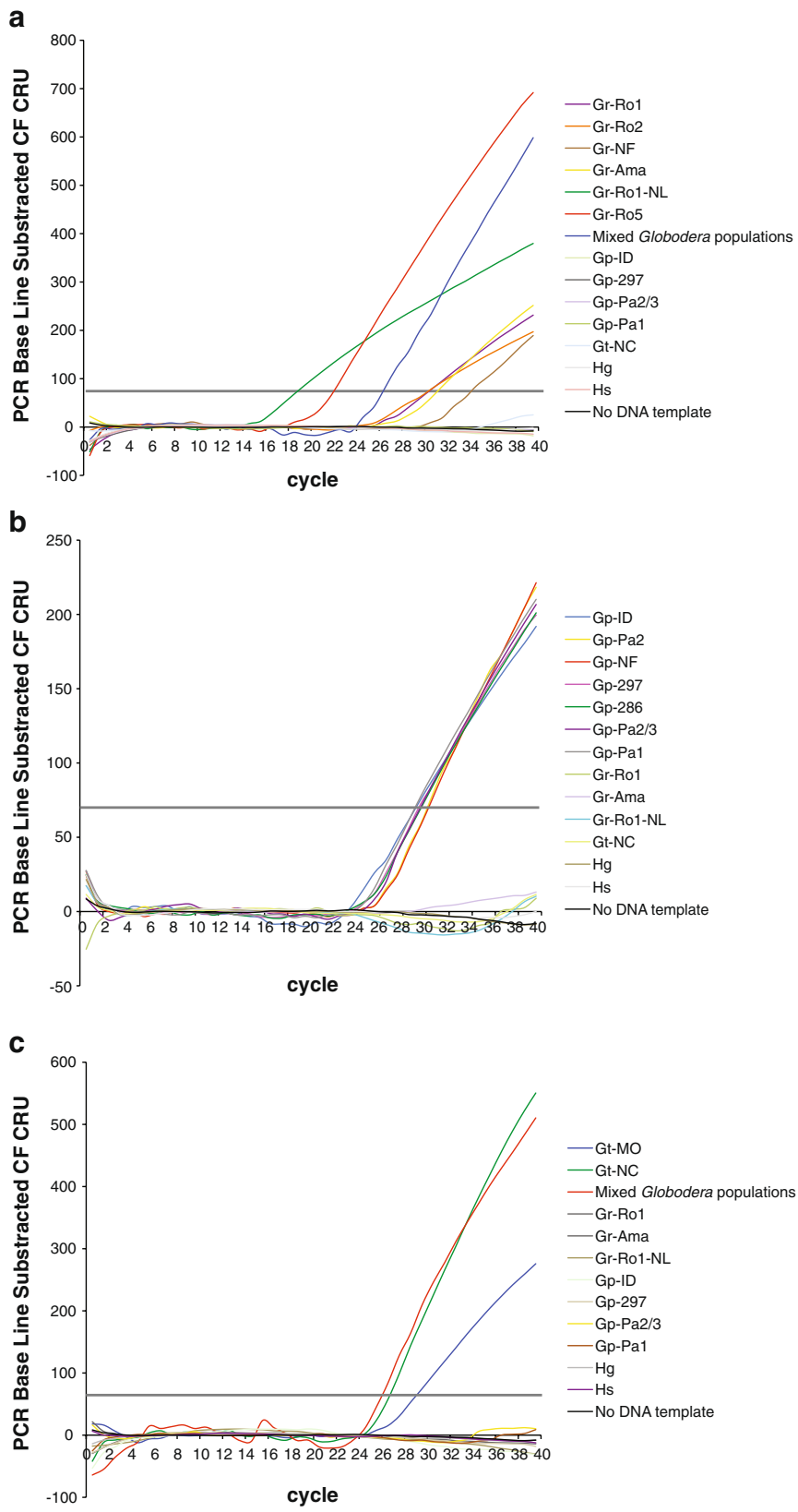

Fig. 4 Amplification plot of TaqMan quantitative real-time PCR assays showing specificity for the identification of *Globodera rostochiensis* **a**, *G. pallida* **b**, and *G. tabacum* **c**. When each species-specific primers-probe set was used in the TaqMan qPCR assays, fluorescence signals were detected only for samples containing DNA from the corresponding nematode species, but not for samples containing no DNA or DNA from other nematode species. Real-time amplified curves shown in different colors represent each tested sample as indicated

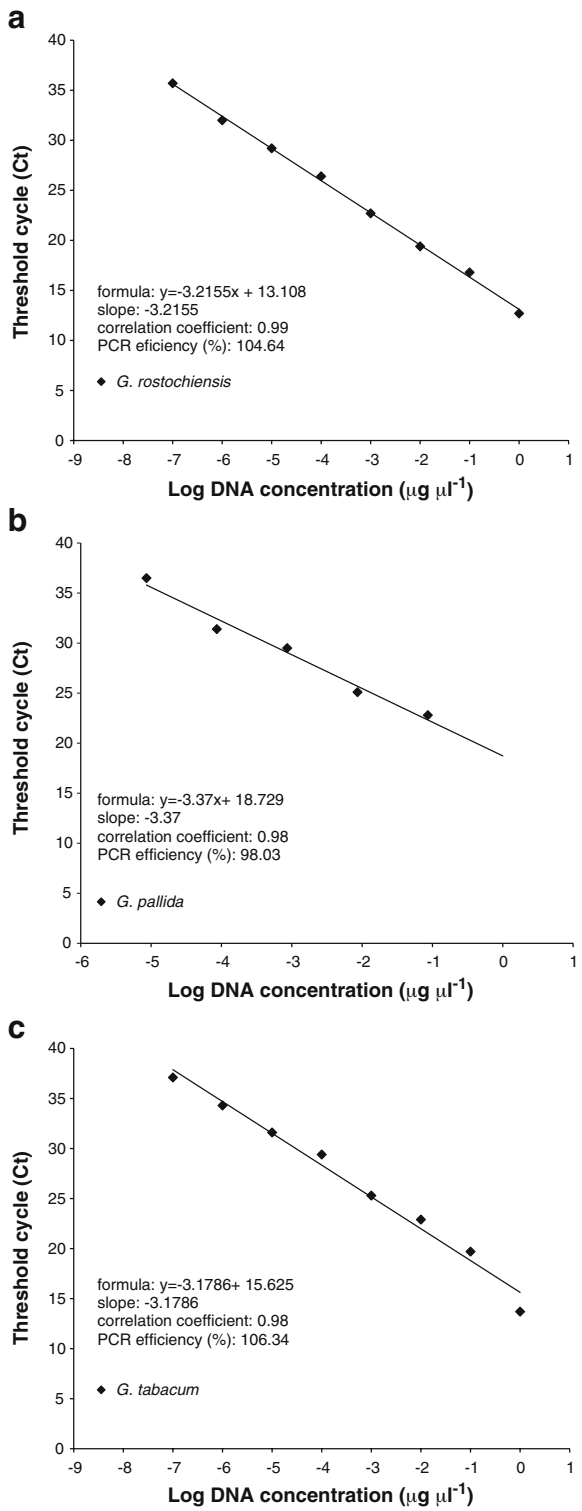

Fig. 5 Standard curves showing threshold cycle (*Ct*) values plotted against the $\log_{10}$ of serially diluted genomic DNA of *Globodera rostochiensis* **a**, *G. pallida* **b**, and *G. tabacum* **c**. Data represent three replicates for each dilution. Similar results were obtained from two independent experiments

populations only when the corresponding nematode-specific primers-probe set was used (Fig. 4b and c). Additionally, each TaqMan qPCR assay gave a positive detection for DNA sample from an artificially-mixed population of three *Globodera* species (Fig. 4). Post-qPCR gel analysis confirmed the amplification of the expected PCR product when a fluorescence signal was detected (data not shown). No signal was detected in TaqMan qPCR assays when samples containing no DNA or DNA from *Heterodera glycines* or *H. schachtii* were used as template (Fig. 4)

Sensitivity of TaqMan qPCR assays

Standard curves were calculated using serial dilutions of genomic DNA from each *Globodera species*. A linear relationship was observed between Ct values and the log concentrations of genomic DNA from *G. rostochiensis* (Fig. 5a), *G. pallida* (Fig. 5b), and *G. tabacum* (Fig. 5c). Correlation coefficients were 0.99, 0.98, and 0.98 for *G. rostochiensis*, *G. pallida*, and *G. tabacum*, respectively, and the efficiencies of the TaqMan qPCR assays were 104.64%, 98.03%, and 106.34% for *G. rostochiensis*, *G. pallida*, and *G. tabacum*, respectively (Fig. 5). The lowest concentrations of DNA samples analyzed that gave a florescence signal were 100 fg μl^{-1} for *G. rostochiensis*, 430 fg μl^{-1} for *G. pallida*, and 100 fg μl^{-1} for *G. tabacum*, therefore, the detection limits of the TaqMan qPCR assays were 100 fg of DNA for *G. rostochiensis* and *G. tabacum* and 430 fg of DNA for *G. pallida*.

Discussion

Alternative pre-mRNA splicing is increasingly recognized as an important mechanism for diversifying the eukaryotic transcriptome and proteome (Blencowe 2006). It has been estimated that more than 70% of human genes and greater than 20% of *Arabidopsis* and rice genes undergo AS (Blencowe 2006; Wang and Brendel 2006). In *C. elegans*, about 10% of genes are estimated to be alternatively spliced (Kim et al. 2007). The chorismate mutase (*CM*) gene from *G.*

rostochiensis (*Gr-cm-1*) was the first reported gene from a plant-parasitic nematode to undergo AS (Lu et al. 2008). The *Gr-cm-1* gene was found to give rise to two splice variants through intron retention, one encoding a functional CM and the other encoding a truncated protein lacking CM activity (Lu et al. 2008). In this study, we cloned a new *CM* gene from *G. tabacum* and further characterized the *CM* gene structure in both *G. tabacum* and *G. pallida*. As expected, we observed that AS of *CM* genes is conserved among the three *Globodera* species, supporting a critical role of AS in regulating *CM* gene function during nematode parasitism (Lu et al. 2008). In addition, based on sequence variations observed among different *Globodera CM* genes, we developed highly specific and sensitive TaqMan qPCR assays for a rapid identification of each *Globodera* species, providing an exciting example of using functional genes other than rDNA sequences as a valuable diagnostic marker for plant-parasitic nematodes.

Chorismate mutase genes from the three *Globodera* species all contain three introns with the second one being the shortest that has a length of 74 bp in *Gr-cm-1* (Lu et al. 2008), 93 bp in *Gp-cm-1*, and 89 bp in *Gt-cm-1*, respectively. AS of *CM* genes through retention of intron 2 was found to be conserved among these *Globodera* species. Intron retention has been found to be associated with weak splice sites and frequently involves short introns (Sakabe and de Souza 2007). The 5' and 3' splice sites of intron 2 are divergent from consensus splice site sequences identified in *C. elegans* (Blumenthal and Steward 1997), suggesting that they might represent suboptimal splice sites. We speculate that these suboptimal splice sites and the short length of intron 2 are important factors contributing to intron retention in *Globodera CM* genes. In addition, putative splicing enhancers predicted in the sequences around intron 2 splice-junctions might also play a role in regulating intron retention. However, in vivo experiments are necessary to confirm these predictions.

The Internal Transcribed Spacer (ITS) regions in ribosomal DNA are widely used as a taxonomic marker for plant-parasitic nematodes (Powers et al. 1997). Various ITS-based molecular methods including recently developed qPCR assays have been proven to be valuable for identifying *Globodera* species (Bates et al. 2002; Madani et al. 2008). Although these qPCR assays offer rapid, specific, and sensitive detection, they utilize limited nucleotide polymorphisms for species differentiation. For example, the EvaGreen-based qPCR assay cannot reliably identify PCN species from samples containing DNA from mixed nematode populations (Madani et al. 2008). The TaqMan qPCR assay developed by Madani et al. (2008) requires the use of expensive LNA-modified primers and probes and its reliability for PCN identification was limited to a narrow range of PCN populations, not including genetically diverse Peruvian populations (Grenier et al. 2001). Although *Globodera CM* sequences are highly similar with each other, a high degree of sequence variations within intron 1, intron 3, and a small region in exon 2 was observed, leading to the development of species-specific TaqMan probes. TaqMan qPCR is generally considered to be more sensitive and offers an added level of specificity compared to SYBRGreen or EvaGreen qPCR. Our developed TaqMan qPCR assays based on using species-specific probes targeting the *CM* gene were confirmed to be extremely specific and sensitive for differentiating *Globodera* species in DNA samples from individual and mixed nematode populations. The sensitivity of the TaqMan qPCR assay developed by Madani et al. (2008) was not reported. Our TaqMan qPCR assays can reliably identify nematode species in samples that contain less than 0.5 pg of nematode DNA. Compared to conventional PCR methods, our TaqMan qPCR assays for PCN identification are more than 1,000 times more sensitive. The high degree of sensitivity and reliability is an added value of our assays, which would be very important especially when nematode material is limited and might permit nematode detection in soil extracts that contain extremely low amounts of nematode DNA. The application of these methods for routine detection in field samples would be valuable for PCN regulatory and quarantine programs. The *CM* gene has been found to be widely expressed in both cyst and root-knot nematode species. The method developed from this study may be adopted for identifying other nematode species, suggesting an important usefulness of the *CM* gene as a diagnostic marker for plant-parasitic nematodes.

Acknowledgements We thank Melissa G. Mitchum for providing TCN and for critical review of this manuscript. We thank Eric Grenier for providing DNA from *G. pallida* populations originated from South Peru, Switzerland and UK, Geert Smant for providing DNA from PCN populations originated

from the Netherlands, Guy Bélair for providing DNA from *G. rostochiensis* populations originated from St.-Amable, Canada, Solke De Boer for providing DNA from PCN populations originated from Newfoundland, Canada, Robert Zemetra for providing *G. pallida* cysts originated from Idaho, U.S.A., and Eric Davis for providing *Heterodera glycines* and a second TCN population from North Carolina, U.S.A. This study was supported in part by funding from USDA-ARS and USDA-APHIS.

References

Bates, J. A., Taylor, E. J. A., Gans, P. T., & Thomas, J. E. (2002). Determination of relative proportions of *Globodera* species in mixed populations of potato cyst nematodes using PCR product melting peak analysis. *Molecular Plant Pathology, 3*, 153–161.

Bélair, G., & Miller, S. (2006). First report of *Globodera tabacum* infecting tobacco plants in Quebec, Canada. *Plant Disease, 90*, 527.

Blencowe, B. J. (2006). Alternative splicing: new insights from global analyses. *Cell, 126*, 37–47.

Blumenthal, T., & Steward, K. (1997). RNA processing and gene structure. In D. L. Riddle, T. Blumenthal, B. J. Meyer, & J. R. Priess (Eds.), *C. ELEGANS II* (pp. 117–145). New York: Cold Spring Harbor Laboratory Press.

Bulman, S. R., & Marshall, J. W. (1997). Differentiation of Australasian potato cyst nematode (PCN) populations using the polymerase chain reaction (PCR). *New Zealand Journal of Crop and Horticultural Science, 25*, 123–129.

Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q., & Krainer, A. R. (2003). ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Research, 31*, 3568–3571.

Davis, E. L., Hussey, R. S., & Baum, T. J. (2004). Getting to the roots of parasitism by nematodes. *Trends in Parasitology, 20*, 134–141.

Doyle, E. A., & Lambert, K. N. (2003). *Meloidogyne javanica* chorismate mutase 1 alters plant cell development. *Molecular Plant-Microbe Interactions, 16*, 123–131.

Dye, B. T., Buvoli, M., Mayer, S. A., Lin, C.-H., & Patton, J. G. (1998). Enhancer elements activate the weak 3' splice site of α-tropomyosin exon 2. *RNA, 4*, 1523–1536.

Grenier, E., Bossis, M., Fouville, D., Renault, L., & Mugniéry, D. (2001). Molecular approaches to the taxonomic position of Peruvian potato cyst nematodes and gene pool similarities in indigenous and important populations of *Globodera*. *Heredity, 86*, 277–290.

Jones, J. T., Furlanetto, C., Bakker, E., Banks, B., Blok, V., Chen, Q., et al. (2003). Characterization of a chorismate mutase from the potato cyst nematode *Globodera pallida*. *Molecular Plant Pathology, 4*, 43–50.

Kabat, J. L., Barberan-Soler, S., McKenna, P., Clawson, H., Farrer, T., & Zahler, A. M. (2006). Intronic alternative splicing regulators identified by comparative genomics in nematodes. *PLoS Computational Biology, 2*, e86.

Kim, E., Magen, A., & Ast, G. (2007). Different levels of alternative splicing among eukaryotes. *Nucleic Acids Research, 35*, 125–131.

Lejeune, F., Cavaloc, Y., & Stevenin, J. (2001). Alternative splicing of intron 3 of the serine/arginine-rich protein 9G8 gene. *Journal of Biological Chemistry, 276*, 7850–7858.

Lu, S.-W., Tian, D., Borchardt-Wier, H. B., & Wang, X. (2008). Alternative splicing: A novel mechanism of regulation identified in the chorismate mutase gene of the potato cyst nematode *Globodera rostochiensis*. *Molecular and Biochemical Parasitology, 162*, 1–15.

Madani, M., Ward, L. J., & De Boer, S. H. (2008). Multiplex real-time polymerase chain reaction for identifying potato cyst nematodes, *Globodera pallida* and *Globodera rostochiensis*, and the tobacco cyst nematode, *Globodera tabacum*. *Canadian Journal of Plant Pathology, 30*, 554–564.

Marché, L., Valette, S., Grenier, E., & Mugniéry, D. (2001). Intraspecies DNA polymorphism in the tobacco cyst-nematode complex (*Globodera tabacum*) using AFLP. *Genome, 44*, 941–946.

Mercado, P. A., Ayala, Y. M., Romano, M., Buratti, E., & Baralle, F. E. (2005). Depletion of TDP 43 overrides the need for exonic and intronic splicing enhancers in the human apoA-II gene. *Nucleic Acids Research, 33*, 6000–6010.

Nielsen, H., Engelbrecht, J., Brunak, S., & von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering, 10*, 1–6.

Powers, T. O., Todd, T. C., Burnell, A. M., Murray, P. C. B., Fleming, C. C., Szalanski, A. L., et al. (1997). The rDNA internal transcribed spacer region as a taxonomic marker for nematodes. *Journal of Nematology, 29*, 441–450.

Sakabe, N. J., & de Souza, S. J. (2007). Sequence features responsible for intron retention in human. *BMC Genomics, 8*, 59.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (Eds.). (1989). *Molecular cloning: A laboratory manual*. New York: Cold Spring Harbor Laboratory Press.

Skantar, A. M., Handoo, Z. A., Carta, L. K., & Chitwood, D. J. (2007). Morphological and molecular identification of *Globodera pallida* associated with potato in Idaho. *Journal of Nematology, 39*, 133–144.

Soller, M. (2006). Pre-messenger RNA processing and its regulation: a genomic perspective. *Cellular and Molecular Life Sciences, 63*, 796–819.

Sun, F., Miller, S., Wood, S., & Côté, M.-J. (2007). Occurance of potato cyst nematode, *Globodera rostochiensis*, on potato in the Saint-Amable region, Quebec, Canada. *Plant Disease, 91*, 908.

Syracuse, A. J., Johnson, C. S., Eisenback, J. D., Nessler, C. L., & Smith, E. P. (2004). Intraspecific variability within *Globodera tabacum solanacearum* using random amplified polymorphic DNA. *Journal of Nematology, 36*, 433–439.

Thiéry, M., & Mugniéry, D. (1996). Interspecific rDNA restriction fragment length polymorphism in *Globodera* species, parasites of Solanaceous plants. *Fundamental and Applied Nematology, 19*, 471–479.

Voelker, R. B., & Berglund, A. J. (2010). A comprehensive computational characterization of conserved mammalian intronic sequences reveals conserved motifs associated with constitutive and alternative splicing. *Genome Research, 17*, 1023–1033.

Wang, B. B., & Brendel, V. (2006). Genomewide comparative analysis of alternative splicing in plants. *Proceedings of the National Academy of Sciences in the United States of America, 103*, 7175–7180.

Wang, X., Allen, R., Ding, X., Goellner, M., Maier, T., de Boer, J. M., et al. (2001). Signal peptide-selection of cDNA cloned directly from the esophageal gland cells of the soybean cyst nematode *Heterodera glycines*. *Molecular Plant-Microbe Interactions, 14*, 536–544.

Zahler, A.M. (2005). Alternative splicing in *C. elegans*. The *C. elegans* Research Community. WormBook. doi/10.1895/wormbook.1.31.1. http//www.wormbook.org.