Physiological and Molecular Plant Pathology 84 (2013) 107–114

Contents lists available at ScienceDirect

Physiological and Molecular Plant Pathology

journal homepage: www.elsevier.com/locate/pmpp

Sequence polymorphism of nematode effectors highlights molecular differences among the subspecies of the tobacco cyst nematode complex

Charline Alenda, Aurore Gallot-Legrand, Didier Fouville, Eric Grenier*

INRA, UMR1349 IGEPP, F-35653 Le Rheu, France

ARTICLE INFO

Article history:
Accepted 28 August 2013

Keywords:
Globodera tabacum
Pectate lyase
Expansin
Clavata like ESR
Species complex
Diagnostic

ABSTRACT

Globodera tabacum is a little-known species complex of specialized plant–parasitic nematodes. In this study we investigated for the first time the sequence variability of five effectors: three genes coding for cell wall degrading enzymes (*Pel1*, *Pel2* and *ExpB3*) and two genes coding for CLE peptides (*CLE1* and *CLE4*) among a set of populations representing the different subspecies described. Substantial variability was observed in the sequence dataset which is consistent with the differential pathogenicity of the *G. tabacum* subspecies on a range of *Nicotiana* species. Using these sequence data, we were able (1) to show the presence of a novel CLE-like class in cyst nematodes, (2) to support the existence of most probably three and not four subspecies in this species complex, (3) to reveal that the subspecies *G. tabacum tabacum* presents the highest diversity and (4) to develop a PCR tool that allows a quick and reliable identification of this subspecies among the species complex.

1. Introduction

Tobacco cyst nematodes (TCN) are responsible for significant yield reductions in the USA, France, Spain and Italy tobacco fields [1]. This highly specialized plant–parasitic nematode species is composed of three subspecies, *Globodera tabacum* subsp. *tabacum* [2,3], *G. tabacum* subsp. *virginiae* [3,4], and *G. tabacum* subsp. *solanacearum* [5,3]. All these subspecies display similar morpho-biometric features, can easily cross both ways and reproduce on tobacco, tomato, bittersweet or black nightshades [6,7]. However, they differ in their ability to damage tobacco cultivars and to develop on different Nicotianae species [8,9].

Because of the current banning of the most effective nematicides and the high survival rate of nematode populations as cysts, host resistance represents an attractive mean of control in infested fields. Several sources of resistance were identified in wild genotypes of Nicotianae but few agronomic cultivars are available. To preserve their efficiencies, resistance must be deployed in an integrated approach through time (to slow down the adaptation) and through space (as no occurrence in sympatry of the TCN subspecies have ever been reported). To date, the subspecies distribution of TCN is poorly known. *Globodera tabacum* is recorded in 16 countries [10], but the subspecies status of these populations remains mainly unknown due to the lack of simple, quick and efficient markers for such identification. Subbotin et al. [11] analyzed the ITS of the rDNA but were not able to identify the different subspecies using the polymorphism revealed by this marker. Only phylogenetic analysis conducted on RAPD [12] or AFLP [13] markers allowed the molecular discrimination of the subspecies inside the TCN complex. Considerable intra-species variability was observed using AFLP with the possibility of a fourth subspecies, *G. tabacum* subspecies "azteca" [13].

Proteins secreted during the migratory and sedentary phases of parasitism, here after called effectors, are thought to be prime candidates to evaluate the adaptation potential of nematodes to plant resistance. Little information is published on effectors of *G. tabacum*. Only one cellulase [14], one pectate lyase [15] and one chorismate mutase [16] are reported so far. However, several of the secreted effectors were isolated and characterized in other species of cyst nematodes (reviewed in Ref. [17]). Among the best characterized, plant cell wall degrading enzymes (CWDE) and regulators of root development like the plant-mimetic peptide CLAVATA3/ESR (CLE), were described and appear to be conserved in *Globodera rostochiensis* [18–20], *Heterodera glycines* [21–23] and *Heterodera schachtii* [24,25]. Secreted in early stages of parasitism, CWDE proteins help infective juveniles to penetrate the roots and migrate into the cells. Secreted later during the parasitic life cycle, CLE proteins appear to act as ligand mimics of endogenous plant

* Corresponding author. Tel.: +33 2 2348 5173.
E-mail address: eric.grenier@rennes.inra.fr (E. Grenier).

0885-5765/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.pmpp.2013.08.004

Table 1
Subspecies and populations used in this study. All these populations were used in the PCR-RFLP analysis, while only twelve were used for gene amplification and sequencing.

Subspecies	Population	Origin	Sequencing	PCR-RFLP
G. t. tabacum	GT3	Mexico	Yes	Yes
G. t. tabacum	Connecticut	USA	Yes	Yes
G. t. tabacum	Agen	France	Yes	Yes
G. t. tabacum	Landes	France	Yes	Yes
G. t. tabacum	Fauillet	France	No	Yes
G. t. tabacum	Aiguillon	France	No	Yes
G. t. virginiae	75117	Mexico	Yes	Yes
G. t. virginiae	GV2	USA	Yes	Yes
G. t. virginiae	GV3	USA	Yes	Yes
G. t. solanacearum	75140	Mexico	Yes	Yes
G. t. solanacearum	GS1	USA	Yes	Yes
G. t. solanacearum	GS2	USA	Yes	Yes
G. t. solanacearum	GS3	USA	No	Yes
G. t. "*azteca*"[a]	75128	Mexico	Yes	Yes
G. t. "*azteca*"[a]	75181	Mexico	Yes	Yes

[a] Species inquiranda proposed by Marché et al., 2001.

peptides to trigger a cascade of developmental events controlling root cell development and central to feeding cell formation [26].

Based on the sequence data already available regarding the effectors of cyst nematodes, we applied a candidate gene approach to further investigate the polymorphisms present in the effectors of TCN and to evaluate their use to distinguish the different TCN subspecies. Twelve populations of TCN representative of the different subspecies were used to sequence the orthologs in *G. tabacum* of the housekeeping gene *EF1α*, three CWDE (two pectate lyases, *Pel1* and *Pel2*, and one expansin-like, *ExpB3*) and two genes coding for CLE peptides (*CLE1* and *CLE4*). Using this sequence dataset we 1/ assessed the within-species and within-subspecies diversity, 2/ investigated if the polymorphism revealed could be correlated with the different subspecies, and 3/ developed rapid and cost-effective molecular markers as a tool for TCN subspecies identification.

2. Materials and methods

2.1. Nematode populations and DNA extraction

The origins of the populations used in this study are summarized in Table 1. Diversity was maximized by choosing populations from different geographical origins. The nematodes were reared in a greenhouse on *Nicotiana tabacum* cv. Samsung. Cysts were extracted three months later using a Kort elutriator, dried and stored at 4 °C. For each population, we used as starting material one cyst that contained hundreds of individuals.

After being presoaked in water overnight, individual cysts were crushed in a separate sterile Eppendorf using a Pasteur pipette with a sealed end. Cell membranes were disrupted by addition of 150 μL of a lysis buffer (200 mM tris–HCl, pH 8.5; 200 mM NaCl, 25 mM EDTA : 0.5% SDS) and 75 μL of sodium acetate (3 M, pH 5.2) followed by a 10 min incubation time at 20 °C. Tubes were centrifuged at 12,000 *g* for 5 min to precipitate the cellular debris. The supernatants were transferred in clean microtubes and the DNAs were precipitated using isopropanol. After a last centrifugation at 12,000 *g* for 30 min, DNA pellets were washed with ethanol 70%, dried under vacuum and suspended in 20 μL of Tris–EDTA 1× (1 mM EDTA, 10 mM Tris–HCl) prior to be used.

2.2. Amplification

Five μL of each resuspended DNA were used to amplify the different genes: the housekeeping gene *EF1α*, two genes coding for different pectate lyases (*Pel1* and *Pel2*), one gene coding for an expansin-like protein (*ExpB3*) and two genes coding for CLE peptides (*CLE1* and *CLE4*). The *G. tabacum* orthologs of *Pel1*, *Pel2*, *CLE1* and *CLE4* have been targeted using primers previously designed [15,20, Grenier, unpublished work]. To amplify *ExpB3* and *EF1α*, we aligned the corresponding EST sequences of *Globodera* and *Heterodera* available in GenBank and designed primers using Primer3 software. All the primers sequences are presented in Supplemental Table 1.

Pel1 was amplified in a final volume of 25 μL using primers PEL1-fwd4 and PEL1-rev3 (10 pmoles each), 1.25U Taq Goflexi (Promega), 5 μL Taq buffer (5×), 5 nmoles of each dNTPs and 150 nmoles of MgCl2. PCR conditions were as follows: 96 °C for 1 min, 35 cycles of 96 °C for 20 s, 55 °C for 20 s and 72 °C for 90 s, ended by a final elongation step at 72 °C during 5 min.

Pel2 was amplified in a final volume of 15 μL using primers PEL2-fwd1 and PEL2-rev1 (10 pmoles each), 0.5U Taq Goflexi, 3 μL Taq buffer (5×), 2.5 nmoles of each dNTPs and 37.5 nmoles of MgCl2. PCR conditions were as follows: 96 °C for 5 min, 35 cycles of 96 °C for 60 s, 58 °C for 60 s and 72 °C for 60 s, ended by a final elongation step at 72 °C during 7 min.

CLE1 was amplified in a final volume of 15 μL using primers CLE1-atgF and CLE1-tgaR [19] (6 pmoles each), 0.5 U Taq Goflexi, 3 μL Taq buffer (5×), 2.5 nmoles of each dNTPs and 75 nmoles of MgCl2. PCR conditions were as follows: 96 °C for 3 min, 35 cycles of 96 °C for 20 s, 64 °C for 20 s and 72 °C for 120 s, ended by a final elongation step at 72 °C during 7 min.

CLE4 (in fact *CLE5*) was amplified in a final volume of 15 μL using primers CLE1-atgF and CLE4-rev (10 pmoles each), 0.5 U Taq Goflexi, 3 μL Taq buffer (5×), 2.5 nmoles of each dNTPs and 37.5 nmoles of MgCl2. PCR conditions were as follows: 96 °C for 3 min, 35 cycles of 96 °C for 60 s, 57 °C for 60 s and 72 °C for 120 s, ended by a final elongation step at 72 °C during 7 min.

ExpB3 was amplified in a final volume of 15 μL using primers EXPB3-rev and EXPB3-fwd (10 pmoles each), 0.5 U Taq Goflexi, 2.5 μL Taq buffer (5×), 2.5 nmoles of each dNTPs and 37.5 nmoles of MgCl2. PCR conditions were as follows: 96 °C for 5 min, 35 cycles of 96 °C for 60 s, 65 °C for 60 s and 72 °C for 120 s, ended by a final elongation step at 72 °C during 7 min.

EF1α was amplified in a final volume of 25 μL using primers of EF1αGlobo-fwd and EF1αGlobo-revB (1 μmole each), 0.04 U Taq Goflexi, 5 μL Taq buffer (5×), 0.1 nmoles of each dNTPs and 1mmole of MgCl2. PCR conditions were as follows: 94 °C for 4 min, 40 cycles of 94 °C for 30 s, 66 °C for 30 s and 72 °C for 60 s, ended by a final elongation step at 72 °C during 5 min.

The DNA fragments were separated by electrophoresis on a 1% TBE buffered agarose gel, stained with SYBRsafe and revealed after a 1 h-migration at 120 V under UV light. DNA bands of interest were recovered from the gel prior cloning by cutting off the gel band and eluting the DNA using the MinElude kit (Qiagen) following the manufacturer's instructions.

2.3. Cloning and sequencing

PCR products were ligated into a pSC-A vector using a strataClone kit (Stratagene) and subsequently transformed into 25 μL of competent cells (Strataclone, Solopack) using a heat shock procedure. PCR amplifications using the M13 forward and reverse primers were used to check the sizes of the inserts of the white colonies and when relevant, clones were sequenced by Macrogen Inc (Korea).

2.4. Sequence alignment and evolutionary analysis

For each sequence, the homology with the targeted gene was checked using a blastn analysis (http://blast.ncbi.nlm.nih.gov/Blast.

cgi). In order to define exon/intron borders, sequences were aligned with the corresponding cDNA of *G. rostochiensis* or *Globodera pallida* (AF127915.1 for *Gr-Pel1*, AY094613.1 for *Gr-Pel2*, JN226216.1 for *Gr-ExpB3*, EU386829.1 for *Gr-CLE1*, EU386844.1 for *Gr-CLE4* and BK006062.1 for *Gp-EF1α*) using the MEGA software (version 4 [27]). DnaSP v5.10 [28] was used to compute several evolutionary statistics like the number of synonymous changes and replacements, the nucleotide diversity (π), the Tajima's *D* statistic and the d*N*/d*S* ratio. DnaSP was also used to compute a sliding window analysis of the nucleotide variability along the different genes using a window size of 50 bp and a step size of 20 bp. Condensed trees based on concatenated exonic sequences were constructed using the maximum likelihood method implemented in MEGA. Choice of the model and model parameters used to construct the tree were carried out using Modeltest [29].

2.5. Cleaved amplified polymorphic sequence (CAPS) for TCN subspecies diagnostic

To further evaluate the interest of some effectors polymorphisms, primers (Supplemental Table 1) were designed to amplify the corresponding flanking regions. Amplifications were conducted on DNA extracted from a single cyst of populations whose subspecies identities were previously resolved. PCR were carried out in a final volume of 25 μL using the relevant primers (0.4 μmole each), 1.25 U Taq Goflexi, 5 μL Taq buffer (5×), 0.2 nmoles of each dNTPs and 3 mmole of $MgCl_2$. PCR conditions were as follows: 95 °C for 1 min, 30 cycles of 95 °C for 30 s, 64 °C for 50 s and 72 °C for 60 s, ended by a final elongation step at 72 °C during 5 min. These PCR products were digested by either *Mbo*I, *Pst*I or *Bst*EII enzymes following the instructions of the manufacturer. The restriction fragment length polymorphisms were revealed by electrophoresis on a 1% TBE buffered agarose gel stained with SYBRsafe.

3. Results

3.1. Gene sequencing results and definition of the sequence datasets

We sequenced three clones per population (except for *EF1α* where five clones were sequenced per population) to constitute our sequence dataset for the analysis of the within-subspecies diversity. Redundant sequences were removed and the remaining sequences were aligned with the corresponding cDNA sequences of potato cyst nematodes to identify the probable exon/intron structure of the genes amplified in *G. tabacum*. Sequences displaying early stop codons and/or different exon/intron boundaries were then removed. Details on each gene dataset are provided below.

Pel1 (GenBank accession numbers HQ850192 to HQ850222). The final *Gt-Pel1* dataset was composed of 27 full-length sequences showing an overall mean genetic distance of 0.016 (K2P distance on the cds with pairwise deletion). The *Gt-Pel1* dataset displayed sequences with open reading frame sizes of 1310 to 1392 bp due to indels in four of the six introns identified. Introns positions and exons sizes were conserved among the different populations and subspecies of TCN and were able to encode a putative protein of 261 aa. The *Gt-Pel1* dataset showed an amino acid identity of 93% with the Gr-PEL1 amino acid sequence.

Pel2 (GenBank accession numbers GQ426924 to GQ426946). A total of 16 sequences was kept in the final dataset. The sequences covered nearly the entire open reading frame of the gene (except the first 23 nucleotides on the 5′ end side and the last 14 nucleotides on the 3′ end side). *Gt-Pel2* sequences displayed an open reading frame of 950 to 957 bp with short indels in one of the two introns identified. The overall mean genetic distance observed in this dataset was 0.009 (K2P distance on the cds with pairwise deletion). Introns positions and sizes of the different exons appeared conserved through the populations and subspecies of TCN. The putative proteins encoded by *Gt-Pel2* sequences shared 73% of identity with the amino acid sequence of Gr-PEL2.

ExpB3 (GenBank accession numbers JN226156 to JN226179). A total of 22 sequences was kept in the final dataset. *Gt-ExpB3* sequences displayed an open reading frame of 1469 to 1531 bp due to indels in the five introns identified. The overall mean genetic distance observed in this dataset was 0.015 (K2P distance on the cds with pairwise deletion). Introns positions and sizes of the different exons appeared conserved through the populations and subspecies of TCN. The putative proteins encoded by *Gt-ExpB3* sequences shared 93% of identity with the Gr-EXPB3 amino acid sequence.

CLE1 (GenBank accession numbers KF516512 to KF516538). The final *Gt-CLE1* dataset was composed of 26 full-length sequences showing an overall mean genetic distance of 0.009 (K2P distance on the cds with pairwise deletion). The *Gt-CLE1* sequences displayed open reading frame ranging from 675 to 679 bp due to short indels in one of the two introns identified. Introns positions and exons sizes were conserved among the different populations and subspecies of TCN and were able to encode a putative protein of 162 aa including a duplicated terminal CLE motif. The *Gt-CLE1* dataset showed an amino acid identity of 91% with the Gr-CLE1 amino acid sequence.

CLE4 (GenBank accession numbers KF553924 to KF553945). A total of 22 sequences showing CLE motifs was obtained using the primers designed to amplify *Gr-CLE4*. Depending on the sequence used, the similarity with *Gr-CLE4* ranged from 67 to 79%. An overall mean genetic distance of 0.084 (K2P distance on the cds with pairwise deletion) was observed when using these 22 sequences. Ten different CLE motifs were identified and one to five repetitions of these motifs were observed per sequence through the dataset (Supplemental Fig. 1). Obviously, different copies of CLE-like genes were present in this dataset and to get a suitable dataset we decided to keep only the sequences displaying the same exon/intron structure and the same number of perfect CLE motifs. Using these criteria we finally got to a dataset composed of 13 full-length sequences (six from Gts, five from Gtt, one from Gtv and one from Gta) showing an overall mean genetic distance of 0.005 (K2P distance on the cds with pairwise deletion). This dataset displayed sequences able to encode putative proteins of either 169 or 170 aa due to one short indel (3 bp) in the second exon. Based on the structure and sequence dissimilarities observed between the CLE motifs of these 13 sequences and other *Globodera* CLE motifs (Supplemental Fig. 1) we decided to group them into a separate class that was named *Gt-CLE5* and that most probably corresponds to a new CLE-like family in *Globodera*.

EF1α (GenBank accession numbers KF553946 to KF554001). The final *Gt-EF1α* dataset was composed of 52 partial sequences showing an overall mean genetic distance of 0.011 (K2P distance on the cds with pairwise deletion) and a global amino acid identity of 93% compared to Gp-EF1α. The 624 bp amplified fragment covered three exons that were conserved in size and position among the different populations and subspecies of TCN and that were able to encode a putative protein of 176 aa representing 38% in length of the Gp-EF1α protein.

3.2. Phylogenetic analysis based on exonic data

Maximum likelihood phylogenetic analyses inferred from the concatenated datasets of *Gt-EF1α*, *Gt-Pel1*, *Gt-Pel2*, *Gt-ExpB3*, *Gt-CLE1* and *Gt-CLE5* revealed three groups corresponding to the subspecies Gtt, Gtv and Gts historically defined (Fig. 1A). The two Mexican populations (75181 and 75128) belonging to the putative fourth Gta subspecies [13] were dispatched in the Gtv and Gts

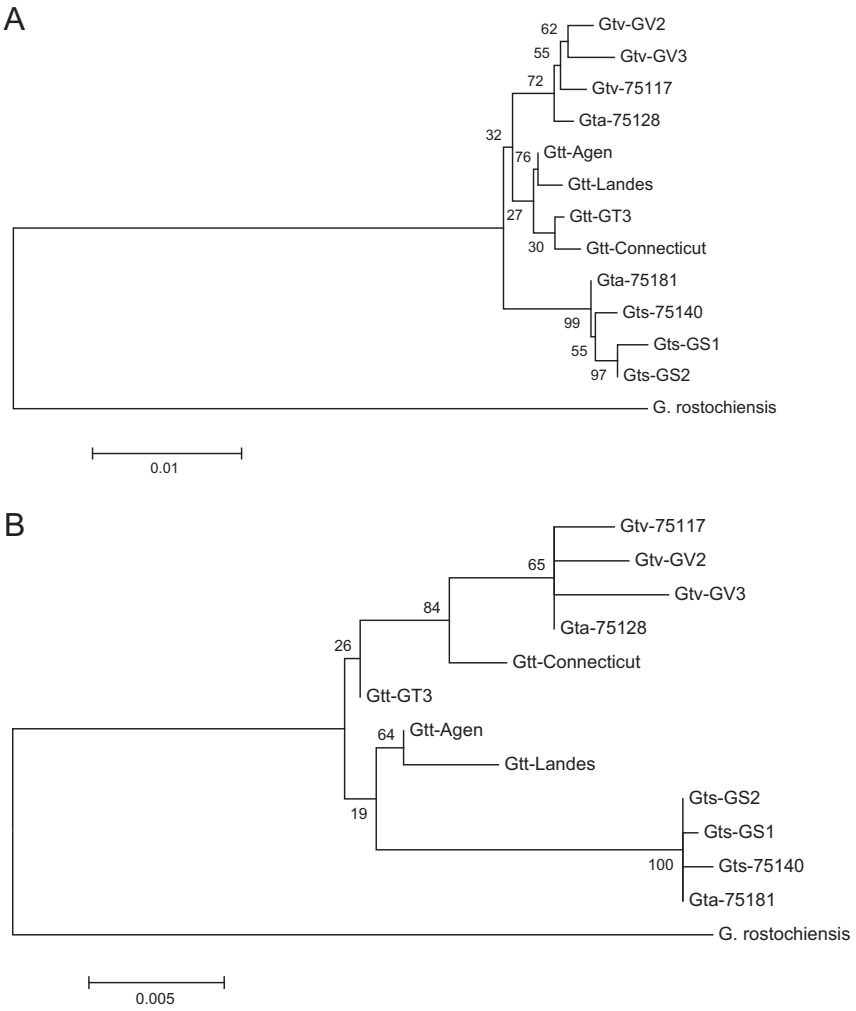

Fig. 1. Phylogenetic analysis of concatenated exons. Analysis were conducted using Maximum Likelihood and 500 bootstraps and the Kimura 2 parameter model in pairwise deletion (K2 + G, with $K = 4$) as defined by modeltest. [A] Tree inferred from the concatenated exons of *Gt-EF1α*, *Gt-Pel1*, *Gt-Pel2*, *Gt-ExpB3*, *Gt-Cle1* and *Gt-Cle5*. [B] Tree inferred from the concatenated exons of *Gt-Pel1*, *Gt-ExpB3* and *Gt-Cle5*.

subspecies. When using concatenated datasets without the three genes that slightly diverged from neutrality (i.e. *Gt-EF1α*, *Gt-Pel2* and *Gt-CLE1*, see Section 3.3) the resolution was worse, the Gtt subspecies appeared no more as a single group but the two Gta populations were still dispatched in the Gtv and Gts subspecies (Fig. 1B). Therefore, we assumed in all the following analysis that populations 75128 and 75181 belong in fact to the Gtv and Gts subspecies, respectively.

3.3. Sequence variability and evolution of G. tabacum effectors

Statistics calculated on the coding regions of the different genes amplified in *G. tabacum* are summarized in Table 2. The five effectors showed from 2% to 10% polymorphic sites equally distributed between synonymous and replacement mutations. We investigated whether evidence for positive or purifying selection pressure could be detected in our data, and carried out evolutionary analyses using the neutrality (Tajima's D) and d*N*/d*S* tests. Values obtained for the Tajima's D statistic provided no highly significant results that allow a clear distinction from neutrality. However, in three of the investigated genes (*Gt-EF1α*, *Gt-Pel2*, and *Gt-CLE1*) significant ($p < 0.05$) negative values were observed that may be indicative of purifying selection. This conclusion was reinforced by the d*N*/d*S* values that were all found <1.

None of the effector targeted displayed a nucleotide diversity significantly different from the one observed in the housekeeping gene ($p > 0.05$) (Fig. 2A). However, *Gt-Pel1* and *Gt-ExpB3* sequences displayed significantly higher nucleotide diversities than the one present in the CLE5 dataset (Fig. 2A). When looking at the distribution of the nucleotide diversity among the different subspecies (Fig. 2B), significant disparities were observed among Gtt, Gtv, and Gts sequences for two of the investigated genes: *Pel1* and *ExpB3*. For three of the four effectors amplified (excluding *Gt-CLE5* as the data were incomplete) the highest levels of nucleotide diversity were observed in the subspecies Gtt while the lowest levels of nucleotide diversity were observed in the subspecies Gts.

3.4. Subspecies specific polymorphisms and development of a Gtt subspecies specific marker

To detect subspecies specific polymorphisms, we computed for each gene a sliding window analysis of the difference between the total nucleotide diversity and the within subspecies nucleotide diversity (i.e. the mean of the three within subspecies diversity). Such differences are helpful to highlight positions where the

Table 2
Polymorphism of the five genes investigated in *Globodera tabacum*: the number of sequences considered within each subspecies is indicated; (*S*) gives the proportion of the polymorphic sites relative to the length of the sequence amplified; (Syn) and (Rep) give respectively the proportion of synonymous mutations and replacements relative to the total number of polymorphic sites; (π) gives the nucleotide diversity and its standard deviation; (d*N*/d*S*) and (D Tajima) statistics give indications on the evolution forces acting on each sequence dataset and deviation from neutrality.

Gene	Nb of Gtt seq	Nb of Gtv seq	Nb of Gts seq	Length (bp)	*S* (%)	π (s.d.)	Syn (%)	Rep (%)	d*N*/d*S*	D Tajima
EF1α	18	15	19	527	11	0.011 (0.001)	46	54	0.08	−1.8 $p < 0.05$
Pel1	8	8	11	785	9.9	0.016 (0.001)	43.6	56.4	0.24	−1.48 ns
Pel2	7	4	5	680	5.3	0.009 (0.001)	41.7	58.3	0.32	−1.86 $p < 0.05$
ExpB3	8	6	8	714	6.7	0.015 (0.001)	43.8	57.2	0.20	−0.76 ns
Cle1	9	8	9	491	7.1	0.009 (0.001)	45.7	54.3	0.31	−1.97 $p < 0.05$
Cle5	5	1	7	513	2.3	0.005 (0.001)	41.7	58.3	0.77	−1.32 ns

sequences are mostly conserved within subspecies while divergent among subspecies. *Gt-Pel2* and *Gt-EF1α* revealed weak differences between the total and the within nucleotide diversities. However some interesting peaks were detected along the sequences of the four other datasets: one in *Gt-CLE5*, two in *Gt-CLE1*, three in *Gt-ExpB3* and five in *Gt-Pel1* (Fig. 3). The nucleotide sequences of the sites related to the highest inflations of the total nucleotide diversity were further analyzed. Polymorphisms tending to be subspecies specific were identified in these four genes: Gtt specific polymorphisms were identified in *Gt-CLE5*, *Gt-CLE1* and *Gt-Pel1*; Gts specific polymorphisms were identified in *Gt-ExpB3*, *Gt-Pel1* and *Gt-CLE1*; and one sole Gtv specific polymorphism was identified in *Gt-Pel1* (Fig. 3). To validate these findings based on sequence analysis, we looked for restriction enzymes with restriction sites matching the polymorphisms identified. We identified three enzymes of interest: *Mbo*I, *Pst*I and *Bst*EII that were used to, respectively, reveal the polymorphisms observed in the *CLE5* "F", *ExpB3* "D" and *CLE1* "E" peaks. PCR primers were designed to amplify the flanking regions of these polymorphisms such a way that the size of the amplified product was short enough to minimize the number of restriction sites subsequently recognized by the enzyme. The efficiency of these three PCR-RFLP markers in separating the different subspecies was tested on a panel of 15 populations of different origin and representative of the *G. tabacum* species complex. As expected, the agarose gel (Fig. 4A) revealed a powerful PCR-RFLP marker specific of all the Gtt populations when using the CAPS marker derived from the *CLE5* "F" peak. However, the usefulness of the CAPS markers derived from the *CLE1* "E" and *ExpB3* "D" peaks to discriminate among the TCN subspecies was hampered by some heterogeneity among the Gtv populations (Fig. 4B and C).

4. Discussion

Few pieces of information were published on effectors of *G. tabacum*. In this study we investigated for the first time the sequence variability of five effectors belonging to different families and involved in different steps of the nematode interaction with its host plant. Substantial variability was observed in the sequence dataset which is consistent with the differential pathogenicity of the investigated subspecies on a range of *Nicotiana* species. Using the data obtained from the investigated genes, we were able (1) to show the presence of a novel CLE-like class in cyst nematodes, (2) to support the existence of most probably three and not four subspecies in the TCN species complex, (3) to reveal that the subspecies Gtt presents the highest diversity and (4) to highlight subspecies polymorphism from which one PCR tool was developed to identify the Gtt subspecies among this complex.

4.1. A large intra-species sequence variability with variations among the copies of a given gene family

Few redundant sequences were obtained in our sequence datasets showing that diversity is high between the individuals within one cyst. As already shown by Philips et al. [30] using subpopulations of *G. pallida* raised from individual cysts, a substantial

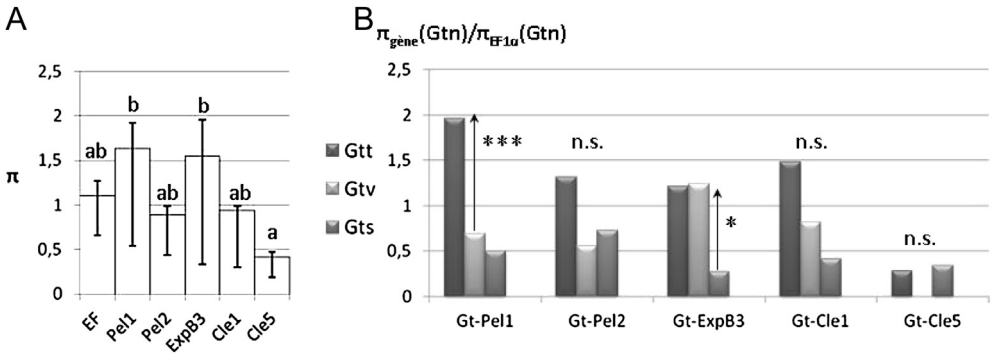

Fig. 2. Nucleotide diversity observed in the effectors. [A] Comparison of the nucleotide diversity observed among the six investigated genes. [B] Comparison of the nucleotide diversities among the TCN subspecies for each of the investigated effectors. For each gene and each TCN subspecies, total nucleotide diversity was corrected by the intra-subspecies diversity occurring in the housekeeping gene (*EF1α*). Statistical significance of the differences observed are indicated above the histograms (n.s. = non significant, *** = $p < 0.01$, * = $p < 0.1$).

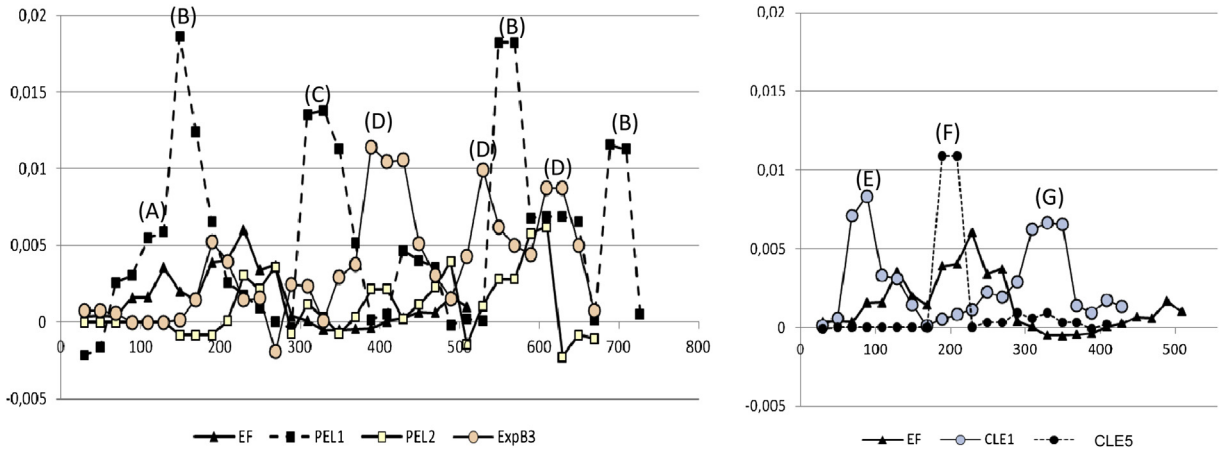

Fig. 3. Difference between nucleotide diversity computed overall TCN populations and within subspecies for the six genes sequenced in *G. tabacum*. Peaks are indicative of inflation of the total nucleotide diversity compared to the mean of the within subspecies nucleotide diversity. Frequencies of the mutations revealed by the inflated peaks in each *G. tabacum* subspecies are detailed below the graphs. Pel1 – peak (A) 70% Gtt, 0% Gtv, 0% Gts ⇒ putative Gtt specific marker. Pel1 – peak (B) 10% Gtt, 0% Gtv, 100% Gts ⇒ putative Gts specific marker. Pel1 – peak (C) 20% Gtt, 100% Gtv, 0% Gts ⇒ putative Gtv specific marker. ExpB3 – peak (D) 0% Gtt, 17% Gtv, 100% Gts ⇒ putative Gts specific marker. Cle1 – peak (E) 0% Gtt, 12.5% Gtv, 78% Gts ⇒ putative Gts specific marker. Cle1 – peak (G) 89% Gtt, 0% Gtv, 11% Gts ⇒ putative Gtt specific marker. Cle5 – peak (F) 100% Gtt, 0% Gtv, 0% Gts ⇒ putative Gtt specific marker.

part of the genetic and phenotypic variability is present among the individuals of a single cyst. One can notice that the mean genetic diversity of the *Pel2* gene in *G. tabacum* (using one cyst per population as starting material) is highly similar to that of the *Pel2* gene in *G. rostochiensis* (using 10 cysts per population as starting material) [15]. We also looked at the minimum number of sequences required

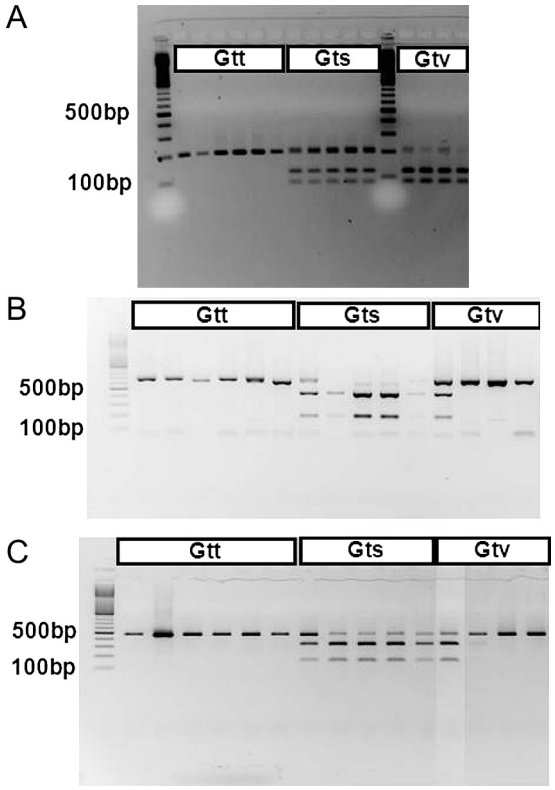

Fig. 4. Evaluation of CAPS markers for the identification of TCN subspecies. [A] PCR-RFLP using the *Mbo*I restriction enzyme to reveal the polymorphism observed in the *Cle5* "F" peak. [B] PCR-RFLP using the *Pst*I restriction enzyme to reveal the polymorphism observed in the *ExpB3* "D" peak. [C] PCR-RFLP using the *Bst*EII restriction enzyme to reveal the polymorphism observed in the *Cle1* "E" peak. In all cases, 15 TCN populations (6 Gtt, 5 Gts and 4 Gtv) of different geographical origins were used to evaluate the specificity and efficiency of the CAPS markers for TCN subspecies identification.

for the correct estimation of the subspecies nucleotide diversity and found that sequencing more than six (sometimes even four) clones of our effectors doesn't improve significantly the measure of this nucleotide diversity (data not shown).

No sign of positive selection was detected in the different genes investigated. The data suggest that the five effectors investigated through this study evolved under purifying selection rather than diversifying selection. However, relatively large intra-species variability was observed among the sequences of each gene. This large intra-species variability appears to be independent of the involvement of the gene in the early or late stages of parasitism. More, this intra-species variability appears to vary among gene copies of a same family (i.e. *Pel1* vs *Pel2* or *CLE1* vs *CLE5*) as much as among genes belonging to different families. The nucleotide diversity of *G. tabacum* effectors ranged from 0.5% to 1.6%, which is a level similar to that observed in other cyst nematode species. In *G. pallida* and *G. rostochiensis*, the *pel2* gene showed 0.5–1.4% nucleotide divergence [15]. In contrast with the results obtained for other expansins isolated in the nematodes *Bursaphelenchus xylophilus* [31] and *G. rostochiensis* [32] for which respectively only 0.1% and 0.4% of divergence were reported, we observed 1.5% of nucleotide diversity in the expansin of *G. tabacum*. Expansins are known to be present as multigenic families in root knot and cyst nematodes and the genome sequence of *Meloidogyne incognita* revealed 20 candidate expansins [33]. They were most probably acquired by nematodes through independent lateral gene transfers followed by gene duplications [34]. It is then tempting to speculate that among this repertoire of cell wall degrading enzyme the different copies can show different levels of sequence divergence depending on the history of duplication of these copies following their acquisition by the nematodes.

4.2. A new class of CLE peptides found in cyst nematodes

CLE peptides secreted by nematodes are known to function as ligand mimics of endogenous plant peptides to trigger a cascade a developmental events central to nematode development [26]. *CLE*-like genes cloned in *G. rostochiensis* encode proteins that usually contain multiple CLE domains. Based on structure and sequence similarities, Gr-CLE were grouped into two classes: Gr-CLE1 and Gr-CLE4. Homologs of the *Gr-CLE1* gene were easily amplified in *G. tabacum*. But, while trying to amplify *Gr-CLE4* in *G. tabacum*, we revealed a complex mixture of divergent sequences, including a

group of 13 sequences appearing as a novel class of CLE proteins in the *Globodera*. This new class was identified as a Gt-CLE5 class. As already observed for Gr-CLE4, Gt-CLE5 contains several members depending on the number or the sequence of the CLE motifs. In contrast to the size variation observed in the coding sequence of Gr-CLE4 [20], the size of the coding sequence of Gt-CLE5 appeared more conserved showing only one indel of 3 bp among the different members. Different families of CLE-like proteins were observed in *G. tabacum*, with some of them showing high variations in term of CLE dodeca sequence and number. It seems therefore that a greater complexity of *CLE*-like genes exists in *G. tabacum* compared to *G. rostochiensis*, a reminder of what was already observed between *CLE*-like genes in *H. glycines* and *H. schachtii* [25,35]. This complexity may also be a reflection of the *G. tabacum* wider host range compared to *G. rostochiensis* as TCN are known to be able to develop also on tomato, pepper, eggplant and on several non cultivated Solanaceae like *Datura stramonium*, *Solanum nigrum* or *Hyoscyamus niger*.

4.3. An update of the subspecies numbering in the G. tabacum complex

Lownsberry and Lownsberry described *G. tabacum* in 1954 [2], while *Globodera virginiae* and *Globodera solanacearum* were described first as distinct species by Miller and Gray in 1968 [4] and 1972 [5], respectively. Stone et al. [3] used canonical analysis based on morphology and morphometrics to place these three entities in a single species complex containing three subspecies. In 2001, Marché et al. [13] studied new populations from Mexico and based on their AFLP results, they proposed a fourth subspecies (*G.t. azteca*) composed of only Mexican populations. The traditional approach for resolving a phylogeny has been to collect data from a single gene across multiple species and populations. Although this approach has been quite useful, it rests upon the assumption that the gene tree will reflect the species tree. Recognition of the fact that gene trees may not accurately reflect the species tree has led researchers to collect data from multiple genes in order to sample independent histories [36–38]. Thanks to the absence of strong divergence from neutrality observed in *G. tabacum*, we were able to improve the phylogenetic analysis after concatenation of the six investigated genes. However, caution must be still taken when using such genes association as effectors usually belong to large gene families where substantial recombination may occur and confound the phylogenetic analysis. The classification of the three groups of TCN obtained in this study is equivalent to that commonly recognized and found by Thierry et al. [12]. Our results do not support the existence of a fourth subspecies as all the Mexican populations studied here cluster either in the Gtv or Gts subspecies. One can argue that may be the slow rate of evolution of nuclear DNA was not enough informative compared to AFLPs and has hampered the identification of the fourth TCN subspecies. However, previous studies showed that nuclear exons can provide robust phylogenetic results and may be still useful even for relatively recent divergences (<10 Myr) [39]. As the genus *Nicotiana* is thought to be between 75 and 100 Myr [40] and as molecular clocks proposed for *Globodera* suggest that potato cyst nematode species could have emerged as early as 15–18 Myr [41], we can assume that TCN subspecies appear relatively recently (i.e. <15 Myr).

4.4. Amount of nucleotide diversity in the different TCN subspecies and development of a Gtt subspecies specific marker

While no significant difference was observed among the nucleotide diversities of the six investigated genes, we observed significant differences among the three TCN subspecies. The Gtt subspecies appeared often as the most diverse subspecies compared to Gtv and Gts. It is unclear at this stage if the higher nucleotide diversity observed in Gtt is an intrinsic feature of this subspecies or is rather related to the presence in our dataset of European populations of Gtt while only populations from North and Central Americas were used for Gtv and Gts. However, as Gtt is known to be the sole TCN subspecies that succeed to establish in countries outside the Americas (and especially in Europe), it is tempting to assume that this particular ability of the Gtt may have been prompted by a wider reservoir of nucleotide diversity present in Gtt nuclear genes. Several works showed that nematode effectors can also represent avirulence genes that will control the compatibility of the interaction with plant resistance genes [42–44]. It is therefore possible that Gtt populations present a higher potential for adaptation to tobacco resistance than Gtv or Gts populations.

Many DNA based molecular diagnostics were developed for application in nematology (see review in Ref. [45]) which are routine, rapid and robust. PCR based molecular diagnostics demonstrate the progress made in this area, offering sensitivity, accuracy and confidence. A major achievement was obtained following the demonstration that a PCR based method could distinguish closely related species [46]. Hence, we tried to discriminate the subspecies of TCN using the sequence data obtained on the *G. tabacum* effectors. Three CAPS markers were designed and evaluated against a set of 15 TCN populations from geographically disparate locations. One Gtt specific marker was obtained using the CAPS marker derived from the *CLE5* "F" peak. This peak stood out from the others as the corresponding polymorphism consisted in an indel rather than a nucleotide substitution. CAPS markers derived from the *ExpB3* "D" and *CLE1* "E" peaks corresponded both to nucleotide substitutions and failed both to be subspecies specific. Our results demonstrate that even substitutions present at a level of 12.5% (as observed in the Gtv for the *CLE1* CAPS marker) is enough to hamper a reliable diagnostic of particular cysts because of partial digestions.

5. Conclusions

Our results showed that effectors are suitable for studying the within species variability but that developing reliable diagnostic markers from these nuclear DNA genes can be hampered by their high level of nucleotide diversity. Despite these limitations, we were able to develop a Gtt specific marker to investigate the presence of this TCN subspecies in fields. Further investigations are now needed (1) to confirm the efficiency of the Gtt diagnostic marker using several cysts per population and (2) to check the absence of cross-hybridization of this marker with other cyst nematode species. Such molecular tool will be of paramount importance to use the appropriate resistance knowing the subspecies status of the TCN populations in place.

Acknowledgment

This work was funded by INRA (Institut National de la Recherche Agronomique), ANITTA (Association Nationale Interprofessionnelle et Technique du Tabac), Imperial Tobacco – ALTADIS and the ARN (Association pour la Recherche sur les Nicotianées) under contract n°3582 "Variability of tobacco cyst nematodes and efficiency of different Nicotiana resistances".

Appendix A. Supplementary material

Supplementary material related to this article can be found at http://dx.doi.org/10.1016/j.pmpp.2013.08.004.

References

[1] La Mondia JA. Shade tobacco yield loss and *Globodera tabacum tabacum* population changes in relation to initial nematode density. J Nematol 1995;27:114–9.

[2] Lownsberry BF, Lownsberry JW. *Heterodera tabacum* new species, a parasite of Solanaceaous plants in Connecticut. Proc Helminthol Soc Wash 1954;21:42–7.

[3] Stone AR, Platt HM, Khalil LF. Three approaches to the status of a species complex, with a revision of some species of *Globodera*. In: Stone AR, Platt HM, Khalil LF, editors. Concepts in nematode systematics. London: Academic Press; 1983. p. 221–3.

[4] Miller LI, Gray BJ. Horsenettle cyst nematode, *Heterodera virginiae* n.sp., a parasite of solanaceous plants. Nematologica 1968;14:535–43.

[5] Miller LI, Gray BJ. *Heterodera solanacearum* n. sp., a parasite of solanaceaous plants. Nematologica 1972;18:404–13.

[6] Motta M, Eisenback J. Morphology of second-stages juveniles and males of *Globodera tabacum tabacum, G. t. virginiae*, and *G. t. solanacearum* (Nemata: Heteroderinae). J Nematol 1993;25:27–33.

[7] Mugniéry D, Bossis M, Pierre JS. Hybridations entre *Globodera rostochiensis* (Wollenweber), *G. pallida* (Stone), *G. virginiae* (Miller and Gray), *G. solanacearum* (Miller and Gray), et *G.* "*mexicana*" (Campos-Vela). Description et devenir des hybrides. Fundam Appl Nematol 1992;15:375–82.

[8] Herrero S, Rufty RC, Barker KR. Evaluation of tobacco germ plasm for resistance to the tobacco cyst nematode, *Globodera tabacum solanacearum*. Plant Dis 1996;80:61–5.

[9] La Mondia JA. Genetics of tobacco resistance to *Globodera tabacum tabacum*. Plant Dis 1991;75:453–4.

[10] CAB International. *Globodera tabacum*. Distribution maps of plant diseases; 2004. Map. 929.

[11] Subbotin SA, Cid Del Prado Vera I, Mundo-Ocampo M, Baldwin JG. Identification, phylogeny and phylogeography of circumfenestrate cyst nematodes (Nematoda: Heteroderidae) as inferred from analysis of ITS-rDNA. Nematology 2011;13:805–24.

[12] Thiéry M, Fouville D, Mugniéry D. Intra and interspecific variability in *Globodera*, parasites of Solanaceous plants, revealed by random amplified polymorphic DNA (RAPD) and correlation with biological features. Fundam Appl Nematol 1997;20:495–504.

[13] Marché L, Valette S, Grenier E, Mugniéry D. Intra-species DNA polymorphism in the tobacco cyst-nematode complex (*Globodera tabacum*) using AFLP. Genome 2001;44:941–6.

[14] Goellner M, Smant G, De Boer JM, Baum TJ, Davis EL. Isolation of beta-1,4-endoglucanase genes from *Globodera tabacum* and their expression during parasitism. J Nematol 2000;32:154–65.

[15] Stare BG, Fouville D, Sirca S, Gallot A, Urek G, Grenier E. Molecular variability and evolution of the pectate lyase (*pel-2*) parasitism gene in cyst nematodes parasitizing different solanaceous plants. J Mol Evol 2011;72:169–81.

[16] Yu H, Chronis D, Lu SW, Wang XH. Chorismate mutase: an alternatively spliced parasitism gene and a diagnostic marker for three important *Globodera* nematode species. Eur J Plant Pathol 2011;129:89–102.

[17] Haegeman A, Mantelin S, Jones JT, Gheysen G. Functional roles of effectors of plant–parasitic nematodes. Gene 2012;492:19–31.

[18] Rehman S, Butterbach P, Popeijus H, Overmars H, Davis EL, Jones JT, et al. Identification and characterization of the most abundant cellulases in stylet secretions from *Globodera rostochiensis*. Phytopathology 2009;99:194–202.

[19] Kudla U, Qin L, Milac A, Kielak A, Maissen C, Overmars H, et al. Origin, distribution and 3D-modeling of Gr-EXPB1, an expansin from the potato cyst nematode *Globodera rostochiensis*. FEBS 2005;579:2451–7.

[20] Lu SW, Chen S, Wang J, Yu H, Chronis D, Mitchum MG, et al. Structural and functional diversity of CLAVATA3/ESR (CLE)-Like genes from the potato cyst nematode *Globodera rostochiensis*. Mol Plant Microbe Interact 2009;22:1128–42.

[21] De Boer JM, McDermott JP, Davis EL, Hussey RS, Popeijus H, Smant G, et al. Cloning of a putative pectate lyase gene expressed in the subventral esophageal glands of *Heterodera glycines*. J Nematol 2002;34:9–11.

[22] Gao BL, Allen R, Maier T, Davis EL, Baum TJ, Hussey RS. Identification of a new beta-1,4-endoglucanase gene expressed in the esophageal subventral gland cells of *Heterodera glycines*. J Nematol 2002;34:12–5.

[23] Wang XH, Mitchum MG, Gao BL, Li CY, Diab H, Baum TJ, et al. A parasitism gene from a plant–parasitic nematode with function similar to CLAVATA3/ESR (CLE) of *Arabidopsis thaliana*. Mol Plant Pathol 2005;6:187–91.

[24] Vanholme B, Van Thuyne W, Vanhouteghem KA, De Meutter J, Cannoot B, Gheysen G. Molecular characterization and functional importance of pectate lyase secreted by the cyst nematode *Heterodera schachtii*. Mol Plant Pathol 2007;8:267–78.

[25] Wang JY, Replogle A, Hussey R, Baum T, Wang XH, Davis EL, et al. Identification of potential host plant mimics of CLAVATA3/ESR (CLE)-like peptides from the plant–parasitic nematode *Heterodera schachtii*. Mol Plant Pathol 2011;12:177–86.

[26] Mitchum MG, Wang XH, Wang JY, Davis EL. Role of nematode peptides and other small molecules in plant parasitism. Annu Rev Phytopathol 2012;50:175–95.

[27] Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 2007;24:1596–9.

[28] Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 2009;25:1451–2.

[29] Posada D. ModelTest Server: a web-based tool for the statistical selection of models of nucleotide substitution online. Nucleic Acids Res 2006;34:W700–3.

[30] Phillips MS, Blok VC, Ploeg A, Harrower BE. Studies on an artificially fragmented population of potato cyst nematode *Globodera pallida*. Nematologica 1998;44:655–66.

[31] Kikuchi T, Hongmei L, Karim N, Kennedy MW, Moens M, Jones JT. Identification of putative expansin-like genes from the pine wood nematode, *Bursaphelenchus xylophilus*, and evolution of the expansin gene family within the Nematoda. Nematology 2009;11:355–64.

[32] Stare BG, Lamovsek J, Sirca S, Urek G. Assessment of sequence variability in putative parasitism factor, expansin (*expB2*) from diverse populations of potato cyst nematode *Globodera rostochiensis*. Physiol Mol Plant Pathol 2012;79:49–54.

[33] Abad P, Gouzy J, Aury JM, Castagnone-Sereno P, Danchin EGJ, Deleury E, et al. Genome sequence of the metazoan plant–parasitic nematode *Meloidogyne incognita*. Nat Biotechnol 2008;26:909–15.

[34] Danchin EGJ, Rosso MN, Vieira P, de Almeida-Engler J, Coutinho PM, Henrissat B, et al. Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. Proc Natl Acad Sci USA 2010;107:17651–6.

[35] Wang JY, Lee C, Replogle A, Joshi S, Korkin D, Hussey R, et al. Dual roles for the variable domain in protein trafficking and host-specific recognition of *Heterodera glycines* CLE effector proteins. New Phytol 2010;187:1003–17.

[36] Leache AD. Species trees for spiny lizards (Genus *Sceloporus*): Identifying points of concordance and conflict between nuclear and mitochondrial data. Mol Phylogen Evol 2010;54:162–71.

[37] Camargo A, Avila LJ, Morando M, Sites JW. Accuracy and precision of species trees: effects of locus, individual, and base pair sampling on inference of species trees in lizards of the *Liolaemus darwinii* group (Squamata, Liolaemidae). Syst Biol 2012;61:272–88.

[38] Lee JY, Joseph L, Edwards SV. A species tree for the Australo-Papuan fairywrens and allies (Aves: Maluridae). Syst Biol 2012;61:253–71.

[39] Steppan SJ, Adkins RM, Spinks PQ, Hale C. Multigene phylogeny of the Old World mice, Murinae, reveals distinct geographic lineages and the declining utility of mitochondrial genes compared to nuclear genes. Mol Phylogen Evol 2005;37:370–88.

[40] Uchiyama H, Chen K, Wildman SG. Polypeptide composition of fraction I protein as an aid in the study of plant evolution. Stadler Symp 1977;9:83–99.

[41] Picard D, Sempere T, Plantard O. Direction and timing of uplift propagation in the Peruvian Andes deduced from molecular phylogenetics of highland biotaxa. Earth Planet Sci Lett 2008;271:326–36.

[42] Bekal S, Niblack TL, Lambert KN. A chorismate mutase from the soybean cyst nematode *Heterodera glycines* shows polymorphisms that correlate with virulence. Mol Plant Microbe Interact 2003;16:439–46.

[43] Sacco MA, Koropacka K, Grenier E, Jaubert MJ, Blanchard A, Goverse A, et al. The cyst nematode SPRYSEC protein RBP-1 Elicits *Gpa2*- and *RanGAP2*-dependent plant cell death. PLoS Pathog 2009;5:e1000564.

[44] Lozano-Torres JL, Wilbers RHP, Gawronski P, Boshoven JC, Finkers-Tomczak A, Cordewener JHG, et al. Dual disease resistance mediated by the immune receptor *Cf-2* in tomato requires a common virulence target of a fungus and a nematode. Proc Natl Acad Sci USA 2012;109:10119–24.

[45] Powers T. Nematode molecular diagnostics: from bands to barcodes. Annu Rev Phytopathol 2004;42:367–83.

[46] Powers TO, Harris TS. A polymerase chain reaction method for identification of five major *Meloidogyne* species. J Nematol 1993;25:1–6.