

This Provisional PDF corresponds to the article as it appeared upon acceptance. Fully formatted PDF and full text (HTML) versions will be made available soon.

Rather than by direct acquisition via lateral gene transfer, GHF5 cellulases were passed on from early Pratylenchidae to root-knot and cyst nematodes

BMC Evolutionary Biology 2012, **12**:221 doi:10.1186/1471-2148-12-221

Katarzyna Rybarczyk-Mydłowska (Kasia.Rybarczyk@wur.nl)

Hazel Ruvimbo Maboreke (maborekh@hu-berlin.de)

Hanny van Megen (Hanny.vanmegen@wur.nl)

Sven van den Elsen (Sven.vandenElsen@wur.nl)

Paul Mooyman (Paul.Mooyman@wur.nl)

Geert Smant (Geert.Smant@wur.nl)

Jaap Bakker (Jaap.Bakker@wur.nl)

Johannes Helder (Hans.Helder@wur.nl)

ISSN 1471-2148

Article type Research article

Submission date 11 July 2012

Acceptance date 4 November 2012

Publication date 21 November 2012

Article URL <http://www.biomedcentral.com/1471-2148/12/221>

Like all articles in BMC journals, this peer-reviewed article can be downloaded, printed and distributed freely for any purposes (see copyright notice below).

Articles in BMC journals are listed in PubMed and archived at PubMed Central.

For information about publishing your research in BMC journals or any BioMed Central journal, go to

<http://www.biomedcentral.com/info/authors/>

© 2012 Rybarczyk-Mydłowska *et al.*

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Rather than by direct acquisition via lateral gene transfer, GHF5 cellulases were passed on from early Pratylenchidae to root-knot and cyst nematodes

Katarzyna Rybarczyk-Mydłowska^{1*}

* Corresponding author

Email: Kasia.Rybarczyk@wur.nl

Hazel Ruvimbo Maboreke²

Email: maborekh@hu-berlin.de

Hanny van Megen¹

Email: Hanny.vanmegen@wur.nl

Sven van den Elsen¹

Email: Sven.vandenElsen@wur.nl

Paul Mooyman¹

Email: Paul.Mooyman@wur.nl

Geert Smant¹

Email: Geert.Smant@wur.nl

Jaap Bakker¹

Email: Jaap.Bakker@wur.nl

Johannes Helder^{1*}

* Corresponding author

Email: Hans.Helder@wur.nl

¹ Laboratory of Nematology, Department of Plant Sciences, Wageningen University, Droevendaalsesteeg 1, 6708 PB, Wageningen, The Netherlands

² Mathematisch-Naturwissenschaftliche Fakultät I, Institut für Biologie, Ökologie, Unter den Linden 6, Berlin 10099, Germany

Abstract

Background

Plant parasitic nematodes are unusual Metazoans as they are equipped with genes that allow for symbiont-independent degradation of plant cell walls. Among the cell wall-degrading enzymes, glycoside hydrolase family 5 (GHF5) cellulases are relatively well characterized, especially for high impact parasites such as root-knot and cyst nematodes. Interestingly, ancestors of extant nematodes most likely acquired these GHF5 cellulases from a prokaryote donor by one or multiple lateral gene transfer events. To obtain insight into the origin of

GHF5 cellulases among evolutionary advanced members of the order Tylenchida, cellulase biodiversity data from less distal family members were collected and analyzed.

Results

Single nematodes were used to obtain (partial) genomic sequences of cellulases from representatives of the genera *Meloidogyne*, *Pratylenchus*, *Hirschmanniella* and *Globodera*. Combined Bayesian analysis of ≈ 100 cellulase sequences revealed three types of catalytic domains (A, B, and C). Represented by 84 sequences, type B is numerically dominant, and the overall topology of the catalytic domain type shows remarkable resemblance with trees based on neutral (= pathogenicity-unrelated) small subunit ribosomal DNA sequences. Bayesian analysis further suggested a sister relationship between the lesion nematode *Pratylenchus thornei* and all type B cellulases from root-knot nematodes. Yet, the relationship between the three catalytic domain types remained unclear. Superposition of intron data onto the cellulase tree suggests that types B and C are related, and together distinct from type A that is characterized by two unique introns.

Conclusions

All Tylenchida members investigated here harbored one or multiple GHF5 cellulases. Three types of catalytic domains are distinguished, and the presence of at least two types is relatively common among plant parasitic Tylenchida. Analysis of coding sequences of cellulases suggests that root-knot and cyst nematodes did not acquire this gene directly by lateral genes transfer. More likely, these genes were passed on by ancestors of a family nowadays known as the Pratylenchidae.

Keywords

Lateral gene transfer, Cellulase, Nematodes, Plant parasitism

Background

Any movement of genetic information, other than by vertical transmission from parents to their offspring via conventional reproduction, is defined as horizontal or lateral gene transfer (HGT or LGT). Although LGT occurs frequently among members of the Archaea and Bacteria, there are only a few probable cases of LGT between prokaryotes and multicellular eukaryotes that have resulted in new functional genes for the recipient. Likely cases of LGT in which the eukaryote is acting as a donor have been described for two mosquito species, *Aedes aegypti* and *Anopheles gambiae* [1]. The transfer of a gene related to malaria sporozoite invasion from mosquito to its endosymbiotic bacterium *Wolbachia pipientis* was demonstrated by Woolfit et al. (2008, [1]). This gene showed substantial divergence, and the level of expression suggested it to be functional in the new prokaryote host. Inter-domain gene transfers can also happen in the reverse way. The pea aphid *Acyrtosiphon pisum* probably acquired two genes from bacteria by LGT [1]. These laterally transferred genes are expressed in the bacteriocytes, and they contribute to the maintenance of *Buchnera aphidicola*, the aphid's primary symbiont. Donors from multiple domains (bacteria, fungi and plants) are thought to be implicated in the acquisition of at least ten protein-coding sequences by the bdelloid rotifer *Adineta vaga* [1]. A subset of these genes were transcribed and

correctly spliced. Interestingly, the authors hypothesized that LGT could be facilitated by mechanisms underlying the desiccation tolerance of this rotifer.

The lateral gene transfer of prokaryotic genes has presumably also played a key role in the evolution of plant parasitism in nematodes. Plant cells are protected by a cell wall, and penetration of this wall is a prerequisite to reach the cytosol. Potato and a soybean cyst nematode (*Globodera rostochiensis* and *Heterodera glycines*) were the first animals shown to harbor symbiont-independent cellulases [2]. These cellulases are classified as members of the glycoside hydrolase family 5 (GHF5). The nematode cellulases appeared to be most similar to bacterial cellulases. In an editorial comment Keen and Roberts [3] suggested that lateral gene transfer may drive the mobility of “pathogenicity islands” (including cellulases) from one organism to the other. Over the last decade, plant parasitic nematodes were shown to harbor a wide spectrum of cell wall-degrading proteins such as pectate lyases [4], polygalacturonase [5], xylanases [6] and expansins [7]. These genes are expressed during infective life stages, and contribute to nematode’s ability to exploit plants as a food source.

Bacterivory is generally accepted as the ancestral feeding type of nematodes. A longstanding hypothesis suggests that bacterivores gave rise to fungivorous nematodes, and facultative and obligatory plant parasites arose from fungal feeding ancestors [8]. It is conceivable that the evolution of plant parasitism in nematodes was driven by the lateral transfer of genes via ingestion of the donor (soil bacteria) by the recipient (bacterivorous nematodes) [9]. Mechanisms underlying desiccation tolerance could have facilitated the uptake of prokaryotic DNA [10]. A number of nematode species including *Aphelenchus avenae* [11], *Ditylenchus dipsaci* [12], and *Panagrolaimus superbus* [13] can develop into highly drought resistant Dauer larva.

Among nematode genes that could have been acquired via one or multiple HGT events, GHF5 cellulases are best characterized. Recent genome sequencing projects resulted in the identification of large cellulase families in the root-knot nematodes *Meloidogyne incognita* [14,15] and *Meloidogyne hapla* [16]. These are highly derived (distal) species within the family Meloidogynidae, and to identify possible origin(s) of these genes, cellulase sequence information is required from less derived representatives of this family. Recent morphological and molecular studies based on female gonoduct architecture [17] and small subunit ribosomal DNA sequences [17,18] suggest that root-knot nematodes originate from - and constitute a subclade within - the genus *Pratylenchus*. By sequencing cellulase genes from *Pratylenchus spp.* (lesion nematodes) and basal root-knot nematode species - the ones that do not belong to one of the subclades I, II and III as defined in 2002 by Tandingan De Ley *et al.* [19] -, we intended to generate clues to establish the evolutionary relationship between members of the Pratylenchidae genera *Pratylenchus* and *Hirschmanniella*, and basal root-knot nematode species such as *Meloidogyne ichinohei*, *M. mali* and *M. ulmi* [18].

Several models have been proposed about HGT event(s) underlying the acquisition of cellulases by plant parasitic and fungivorous nematodes. So far it is unclear whether the distribution of cellulase-encoding genes among Tylenchida is the result of a single HGT event, followed by early single duplication event as suggested by Kyndt *et al.* [20], or the outcome of multiple HGT events. Comparison of the topologies of phylogenetic trees based on SSU rDNA data (*e.g.* [18,21]) *versus* GHF5 cellulase-based tree might tell us whether the evolution within the Pratylenchidae / Meloidogynidae branch includes one or multiple distinct cellulase lineages. Analysis of 103 paralogs and orthologs of cellulase-encoding gene(s) (fragments) from plant parasitic Tylenchida revealed a major clade with a topology

similar to the one revealed by SSU rDNA, a neutral gene. Moreover, a relatively small, divergent subset of cellulases was found that is probably the result of early substrate specificity-driven diversification. Within the catalytic domain types A and B (too few type C sequences are available to make a statement), the overall topology resembles the topologies revealed by neutral ribosomal DNA sequences, and it is hypothesized that root-knot, cyst and lesion nematodes received their cellulases from more ancient Pratylenchidae or even more basal members of Clade 12 [22], rather than by direct lateral gene transfer.

Methods

Taxon sampling and microscopic identification

Pratylenchus and *Hirschmanniella* species were collected from various habitats throughout The Netherlands and extracted from the soil using standard techniques. Individual nematodes were identified using a light microscope (Zeiss Axioscope) equipped with differential interference contrast optics. *Globodera pallida* specimens originated from a Dutch population named Pa3 – Rookmaker. *Meloidogyne* species were kindly provided by Dr. Gerrit Karssen from the Plant Protection Service of The Netherlands: *M. ichinochei* (propagated on *Iris levigata*; culture C2312; Japan), *M. artiellia* (sampled from a field with *Triticum aestivum*; culture E8067; Syria), *M. ardenensis* (propagated on *Liguster sp.*; Wageningen) and *M. ulmi* (isolated from *Ulmus sp.*; Wageningen).

Nematode lysis

Single nematodes were transferred to a 0.2 mL polymerase chain reaction (PCR) tube containing 25 μ L sterile water. An equal volume of lysis buffer containing 0.2 M NaCl, 0.2 M Tris-HCl (pH 8.0), 1% (v/v) β -mercaptoethanol and 800 μ g/mL proteinase K was added. Lysis took place in a Thermal cycler (MyiQ, Bio-Rad) at 65 °C for 2 h followed by 5 min incubation at 100 °C. The lysate (crude DNA extract) was used immediately or stored at -20°C.

Amplification of cellulase-coding genes from genomic DNA

Based on publicly available cellulase sequences (cDNA and genomic sequences, see Additional file 1: Table S1) from lesion, root-knot, and cyst nematodes, six conserved peptide motives were identified within the catalytic domain, namely PPYGQLS (CD1), LKCNWN (CD2), YVIVDW (ENG1), WCQDV (CD4), FVTEYG (ENG2) and ISYLNWAISD (CD6) (for positioning see Table 1). These regions were used as a starting point for the design of *eng*-specific primers (Table 1). The primary aim was to amplify the longest possible fragment, preferably from CD1 to CD6 (230 amino acids, \approx 700bp of the coding sequence); however, on some occasions, only shorter cellulase fragments could be amplified (Table 2).

Table 1 Overview of PCR primers used for cellulase amplification from individual nematodes

A		A	
CD1		CD2	
<u>TATPPPYGQLSVSGTKLVDSSGQPVQLIGNSLFWHQFQAQYWNAETVKALKCNWNANVVRAAVGVDLERGYMSDP</u>			
ENG1			
TTAYNQAVAVIEAAISQGLYVIVDWHSHESHVDKAIEFFTKIAKAYGSYPHVL YETFNEPLQGVSWTDILVPYHKKVIAAI			
CD4		ENG2	
RALDSKNVILGTPTWCQDV DIASQNP IKEYKNLMYTFHFYAATHFVNGLGAKLQTAINNGLPIFVTEYGTCSADGNGNI			
CD6			
DTNSISSWWSLMDNLKISYLNWAISDKSETCSALKPGTPAANVGVSSSWTTSGNMVADHDKKKSTGVSCS			
B Primer sequence 5'→3' *		Primer sequence 5'→3' *	
Region PPYGQLS (CD1):		Region WCQDV (CD4):	
CD1aF	ccIccItacggIcaattg tc	CD4aR	tccacRtcc ctggg acca
CD1bF	ccIccItatggIcaattg tc	CD4cR	tccacAtc t gg ca cca
CD1cF	ccIccItatggIcaatt atc		
CD1PraFa	ccg ccg tatggg caa	Region FVTEYG (ENG2):	
CD1PraFb	cc tccct atgg caa	ENG2 (see <i>e.g.</i> [20])	gtIccRtaYTcIgtIacRaa
CD1PraFc	cg ccg tatggg caa		
CD1MelF	ctcc at atggg caatt atctgt	Region ISYLNWAISD (CD6):	
Region LKCNWN (CD2):		CD6PraFb	tctcctacatcaactgggc
CD2aF	ct caaatgcaatt gga acKc	CD6aR	gcccag ttggcg ta Igaga
CD2bF	ct caaatgcaatt gga atKc	CD6bR	gocca I ttg gcRtaI gaa
CD2cF	ct t aaatg caI ttgga atKc	CD6cR	gocca I ttg aI gta Mgaaa
CD2dF	ct t aaatg ctI ttgga atKc	CD6dR	gcccag ttgaY gta Igaga
Region YVIVDW (ENG1):		CDGp8R	gcccag ttgagg ta cgaa
ENG1 (see <i>e.g.</i> [20])	taYgtIatcgtIgaYtggca	CD6PraRa	cccag ttggcg tagga
ENG1R	tgccaRtcIacgatIacRta	CD6MelR	tgtttgagatagcccag ttg

* I=inosine; K= g or t, M= a or c, R= a or g, Y= c or t. In bold: discriminative nucleotide position

Conserved amino acid motives in GHF5 cellulases from plant parasitic nematodes residing in nematode Clade 12 (Holterman *et al.* 2006 [22]). Primer design was based on these motives, and all primers used in this study are listed below. **A.** The backbone sequence given below is derived from the predicted amino acid sequence of the potato cyst nematode (*Globodera rostochiensis*) cellulase *Gr-eng-1* (GenBank)

AF004523, amino acid positions 18 – 324 (mainly catalytic domain). Underlined: part of signal peptide for secretion. **B.** Primer names and primer sequences

Table 2 Overview of GHF5 cellulase sequences generated in this study from plant parasitic nematodes belonging to the superfamily Hoplolaimoidea

Species name	Individual*	Forward primer (1)	Reverse primer (1)	Fragment length (bp)	Fragment length after removal putative introns (bp)	Gene name (2)	GenBank Acc. No.
<i>Meloidogyne ardenensis</i>	1	CD1cF	CD6cR	1228	727	<i>Mard-eng-1</i>	JN052024
<i>Meloidogyne artiellia</i>	1	CD1MeIF	CD6MeIR	936	731	<i>Mart-eng-1</i>	JN052025
<i>Meloidogyne ichinochei</i>	1	CD2aF	CD4aR	673	343	<i>Mic-eng-1</i>	JN052026
<i>M. ichinochei</i>	2	CD2aF	CD4aR	673	343	<i>Mic-eng-2</i>	JN052027
<i>Hirschmanniella gracilis</i>	1	CD1aF	CD6bR	950	728	<i>Hgr-eng-1</i>	JN052061
<i>H. gracilis</i>	1	CD1aF	CD6bR	866	728	<i>Hgr-eng-2</i>	JN052062
<i>H. gracilis</i>	1	CD1aF	CD6bR	950	728	<i>Hgr-eng-3</i>	JN052063
<i>Hirschmanniella loofi</i>	1	CD1bF	ENG1R	282	255	<i>Hl-eng-1</i>	JN052057
<i>H. loofi</i>	1	CD1bF	CD4aR	560	487	<i>Hl-eng-2</i>	JN052058
<i>H. loofi</i>	1	ENG1	CD6aR	983	461	<i>Hl-eng-3</i>	JN052059
<i>H. loofi</i>	1	ENG1	CD6aR	554	431	<i>Hl-eng-4</i>	JN052060
<i>Pratylenchus crenatus</i>	1	CD1PraFa	CD6aR	820	721	<i>Pcr-eng-1</i>	JN052031
<i>P. crenatus</i>	2	CD1PraFb	CD6PraRa	819	723	<i>Pcr-eng-2</i>	JN052030
<i>P. crenatus</i>	3	ENG1F	CD6aR	543	449	<i>Pcr-eng-3</i>	JN052029
<i>Pratylenchus neglectus</i>	1	ENG1	CD6aR	508	452	<i>Pn-eng-1</i>	JN052032
<i>P. neglectus</i>	2	CD1PraFc	CD6PraRa	791	735	<i>Pn-eng-2</i>	JN052033
<i>P. neglectus</i>	2	CD1PraFc	CD6PraRa	789	733	<i>Pn-eng-3</i>	JN052034
<i>Pratylenchus penetrans</i>	1	CD2cF	CD6aR	1514	588	<i>Pp-eng-3</i>	JN052035
<i>P. penetrans</i>	1	CD2cF	CD6aR	695	587	<i>Pp-eng-4</i>	JN052036
<i>P. penetrans</i>	1	CD1bF	CD4aR	739	484	<i>Pp-eng-5</i>	JN052037
<i>P. penetrans</i>	2	CD1PraFb	CD6PraRb	841	733	<i>Pp-eng-6</i>	JN052038
<i>Pratylenchus convalariae</i> (3)	1	CD1PraFb	CD6aR	874	739	<i>Pcon-eng-1</i>	JN052028
<i>Pratylenchus pratensis</i>	1	CD1bF	ENG1R	256	256	<i>Ppr-eng-1</i>	JN052040
<i>P. pratensis</i>	1	CD2bF	CD6dR	1124	592	<i>Ppr-eng-2</i>	JN052039
<i>P. pratensis</i>	1	CD2dF	CD4aR	407	349	<i>Ppr-eng-3</i>	JN052041
<i>P. pratensis</i>	1	CD1aF	CD4aR	489	489	<i>Ppr-eng-4</i>	JN052043
<i>P. pratensis</i>	1	ENG1	CD6aR	507	452	<i>Ppr-eng-5</i>	JN052042
<i>P. pratensis</i>	2	CD1PraFb	CD6PraRa	801	750	<i>Ppr-eng-6</i>	JN052044
<i>Pratylenchus thornei</i>	1	CD2bF	CD6dR	678	587	<i>Pt-eng-1</i>	JN052045
<i>P. thornei</i>	2	CD1PraFc	CD6PraRb	824	733	<i>Pt-eng-2</i>	JN052046
<i>Pratylenchus vulnus</i>	1	CD2cF	CD6aR	639	587	<i>Pv-eng-1</i>	JN052047
<i>P. vulnus</i>	1	CD2bF	CD6dR	1001	614	<i>Pv-eng-2</i>	JN052048
<i>P. vulnus</i>	1	CD2bF	CD6dR	639	587	<i>Pv-eng-3</i>	JN052049
<i>P. vulnus</i>	1	CD2aF	CD4aR	349	349	<i>Pv-eng-4</i>	JN052050
<i>P. vulnus</i>	1	ENG1	CD6aR	958	449	<i>Pv-eng-5</i>	JN052051
<i>P. vulnus</i>	1	ENG1	CD6aR	604	312	<i>Pv-eng-6</i>	JN052052
<i>P. vulnus</i>	1	ENG1	CD6aR	505	452	<i>Pv-eng-7</i>	JN052053
<i>P. vulnus</i>	2	CD1aF	CD6bR	1031	728	<i>Pv-eng-8</i>	JN052054
<i>P. vulnus</i>	3	CD1PraFb	CD6aR	790	733	<i>Pv-eng-9</i>	JN052055
<i>P. vulnus</i>	3	CD1PraFb	CD6PraRa	792	735	<i>Pv-eng-10</i>	JN052056
<i>Globodera pallida</i>	1	CD1aF	CD4cR	1144	481	<i>Gp-eng-1</i>	JN052064
<i>G. pallida</i>	1	CD2aF	CD6cR	852	589	<i>Gp-eng-2</i>	JN052065
<i>G. pallida</i>	1	CD2aF	CD6cR	853	590	<i>Gp-eng-3</i>	JN052066
<i>G. pallida</i>	2	CDGp2F	CD6cR	1561	709	<i>Gp-eng-4</i>	JN052067
<i>G. pallida</i>	2	CDGp2F	CDGp8R	706	706	<i>Gp-eng-5</i>	JN052068

Overview of GHF5 cellulase sequences generated in this study from plant parasitic nematodes belonging to the superfamily Hoplolaimoidea. Single nematodes were used for the amplification of putative cellulase fragments. Occasionally, multiple fragments were amplified from an individual, such as in case of *Pratylenchus neglectus* 2 (*Pn-eng-2* and -3). (1) Primers sequences are given in Table 1. (2) gene names in bold indicate fragments spanning almost the full catalytic domain (CD1 - CD6). (3) Small subunit ribosomal DNA data suggest that *Pratylenchus convalariae* is identical to *P. penetrans* [18]

Due to differences in codon usage within the six conserved amino acid motives mentioned above, numerous primer variants had to be designed and subsequently examined in various combinations. For a quick, first selection of the most effective primer combinations, quantitative PCR was used. For this, 3 μ L of template (single nematode lysate) was mixed with relevant primers (end concentrations for both primers 200 nM), and 12.5 μ L iQ Absolute Sybr Green Fluorescein Cat. CM-225 (Westburg). The total reaction volume was 25 μ L. Thermal cycling took place in the MyiQ thermal cycler (Bio-Rad) under the following conditions: 95 °C for 15 min; followed by 60 cycles at 95 °C for 30 s, 50°C for 1 min and 72 °C for 2 min. In case a possibly applicable amplification signal was produced (criteria: C_t value < 50 cycles, and a melting temperature > 80°C), the amplicon was analyzed on a 1% agarose-gel stained with GelStar (Westburg; 2 μ l/100ml). For those primer combinations that gave rise to amplification of the expected size products, the annealing temperature was optimized using conventional PCR. These reactions were performed in a final volume of 25 μ l and contained 3 μ l of a diluted crude DNA extract, 0.1 μ M of each PCR primer, and a Ready-To-Go PCR bead (Amersham, Little Chalfont, Buckinghamshire, UK). The following PCR profile was used: 95°C for 5 min followed by 60 x (94°C, 30 sec; specific annealing temperature, 1 min; 72°C, 2 min) and 72°C, 10 min.

Cloning, sequencing and sequence alignment

Gel-purified amplification products (Marligen Bioscience, Ijamsville, MD) were cloned into a TOPO TA cloning vector (Invitrogen, Carlsbad, CA) and sequenced using standard procedures. Newly generated sequences were deposited at GenBank under accession numbers listed in Table 2.

Intron positions in the genomic sequences were identified on the basis of information about exon-intron structure of publicly available sequences (all the ones in the Additional file 1: Table S1 harboring at least one intron). The newly obtained nucleotide sequences as well as those derived from GenBank were translated into amino acids and aligned using the ClustalW algorithm as implemented in BioEdit 5.0.9 [23]. The protein alignment was improved manually and translated back into nucleotides. The final nucleotide alignment consisted of 103 sequences, of which 45 were generated in this study. More than half of these sequences (66 out of 103) span almost the full catalytic domain (from CD1 to CD6).

RNA extraction and cDNA cellulase amplification

In order to support the chosen intron extraction approach, cDNA cellulase sequence for *G. pallida* was synthesized. For this purpose 100 individuals of *G. pallida* were collected into a 0.2 mL PCR tube containing 25 μ L of sterile water and lysed as specified above. The lysate was used immediately for the RNA extraction according to RNeasy Micro kit protocol (Qiagen). Total RNA end concentration was approximately 7ng of RNA/ μ L of water. 3 μ L of the five times diluted RNA was mixed with CDGp2F and CDGp8R primers (Table 2; end concentrations for both primers 200 nM), 20 units of RNase inhibitor (Invitrogen) and the components of the SuperScript III One-Step RT-PCR with Platinum *Taq* kit (12 μ L of 2X reaction Mix and 2 μ L of the SuperScriptTM III RT/ Platinum *Taq* Mix). This reaction of 25 μ L in total, was used for the specific cDNA fragment amplification under the following conditions: 60 °C for 30 min, 94 °C for 2 min; followed by 60 cycles at 94 °C for 15 s, 60°C for 30 s and 68 °C for 1 min and finished with one cycle of 68 °C for 5 min. As a result of this experiment the Gp-eng-5 sequence was acquired.

Phylogenetic analysis

The Bayesian phylogeny was constructed with the program MrBayes 3.1.2 using a site-specific model. Data were partitioned by codon position and gamma distribution of rate variation with a proportion of invariable sites was used. Four independent runs were set with 4 Markov chains per run. The program was run for 5 million generations. Stabilization of the likelihood and parameters were checked with the program Tracer v1.4 [24] and the burnin was defined as 120,000 generations. For the construction of the maximum likelihood of the cellulase tree the RAxML-HPC BlackBox program [25] available at the CIPRES Science Gateway V. 3.1 was used. The program FindModel (<http://hcv.lanl.gov/content/sequence/findmodel/findmodel.html> from the HCV sequence database) was used to determine the best phylogenetic model, and the following parameters were applied: estimated proportion of invariable sites (GTRGAMMA + I). To find best tree using maximum likelihood search, bootstrapping halted automatically and printed branch lengths. Cellulases from *Aphelenchus avenae*, a predominantly fungivorous species, were used as outgroup as this species does not belong to the Tylenchida (all other species used in this study belong to this order), and - on the basis of SSU rDNA data - resides at the very base of Cade 12.

Results and discussion

Identification of new cellulase genes in nematodes

Genomic DNA from individuals from seven *Pratylenchus* (Pratylenchidae), two *Hirschmanniella* (Pratylenchidae), three basal *Meloidogyne* (Meloidogynidae) and one *Globodera* (Heteroderidae) species were tested for the presence of GHF5 cellulases. In all of the nematodes we identified at least one GHF5 cellulase gene (Table 2). Recently, it was shown that GHF5 cellulases are not the only type of cellulases present among members of the Tylenchida (Bauters *et al.*, pers. comm.). Except for the very basal *Meloidogyne* species *M. ichinochei* and *Hirschmanniella loofi*, we obtained at least one complete CD1-CD6 (Table 2) sequence for every species under investigation.

Genome sequencing of the distal tropical *Meloidogyne* species *M. incognita* not only revealed the presence of multiple (21) genes encoding putative cellulases, but also showed distinct clusters of GHF5 sequences within a single species [14]. Here we demonstrate that this is not unique for distal *Meloidogyne* species. *Pratylenchus pratensis* harbors multiple cellulase genes, whereas both type A or type B catalytic domains (as coined by [20]; Additional file 2: TableS2) are represented. Similar results were obtained for *Pratylenchus vulnus* (Figure 1). The inventory of all currently available catalytic domain sequences points at a numerical dominance of the type B over type A and C catalytic domains. For the soybean cyst nematode *Heterodera glycines*, the substrate specificities of cellulases with type A, B and C catalytic domains were tested [26]. Hg-ENG5 (Type C) and Hg-ENG6 (Type A) differed greatly from the most abundant type B cellulases: their depolymerizing activity on carboxymethylcellulose was strongly reduced (respectively $\approx 40\%$ and $\approx 20\%$ of the activities of HG-ENG-1 and -4 (both belonging to type B cellulases)), whereas both degrade xylan and crystalline cellulose (Hg-ENG5 showing two fold higher activity than Hg-ENG6). By contrast, the latter two substrates were not significantly degraded by Hg-ENG-1 and -4 [26]. As such differences have been reported for other GHF5 cellulases as well (*e.g.* [27]), we hypothesize that the different types of catalytic domains could point at differences in substrate specificities.

Figure 1 Bayesian tree of GHF5 catalytic domains from members of the nematode order Tylenchida. Genomic and coding sequences (indicated by a yellow box at the base of the relevant branch) from (partial) cellulase catalytic domains were analyzed. Sequences covering the catalytic domain from CD1 to CD6 (as defined in Table 1) are underlined (non underlined sequences are slightly shorter). Identical colors are used for members of the same nematode family. The tree is rooted with genomic cellulase sequences from the fungivorous nematode *Aphelenchus avenae* (infraorder Tylenchomorpha). Posterior probabilities are given next to each node. Orange circles with or without a bright cross are used to indicate the presence or absence of an intron. An orange cross behind a sequence is used to indicate that the generated piece of a sequence was intronless. Intron numbering is essentially according to Kyndt et al. (2008) [20]. Branch length is calculated in MrBayes, and the scale bar below represents branch length (as number of DNA substitutions/site)

Phylogenetic analyses

Bayesian inference-based phylogenetic analysis of 103 coding sequences (of which 45 were generated in this study) for - at least - a major part of the catalytic domain of GHF5 cellulases resulted in the distinction of three major types of catalytic domains (A, B, and C, see Figure 1). This systematics elaborates on the evolutionary model proposed by Kyndt *et al.* (2008) [20]; catalytic domain C was originally presented as a well-supported group of cellulases nested within the type B clade [20].

Among cellulases with a type B catalytic domain (comprising 84 out of the 103 sequences), we observed a topology that shows a remarkable overall similarity with the one presented by Holterman *et al.* (2009) [18] on the basis of a neutral gene, *viz.* SSU rDNA. Also on the basis of type B catalytic domain sequences *M. artiellia* and *M. ichinochei* appear at the base of the family Meloidogynidae, and just as observed on the basis of SSU rDNA sequences the Meloidogynidae appear as an elaborate subclade nested within the genus *Pratylenchus*. Unlike SSU rDNA sequences that gave no clear answer about the nature of the link between the Pratylenchidae and the Meloidogynidae, a robust sister relationship was observed between a specific *Pratylenchus* species, *P. thornei*, and all representatives of the genus *Meloidogyne*.

Ribosomal DNA-based phylogenetic analysis revealed a sister relationship between lesion and root-knot nematodes on the one hand, and cyst nematodes, Hoplolaimidae and Rotylenchulidae on the other (*e.g.* [18]). Remarkably, all catalytic domains from the lesion nematode *P. crenatus*, and subsets from the *P. vulnus* and *P. penetrans* cellulases reside in a sister position with regard to all type B cyst nematode cellulases. Although the number of sequences included is considerably smaller, we observed a similar pattern for the type A catalytic domain. Firstly, a sister relationship was established here between all type A catalytic domain sequences from root-knot nematodes and a cellulase from *P. vulnus*, whereas a similar relationship was observed for a soybean cyst and a reniform nematode cellulase on the one hand, and *P. pratensis* on the other (Figure 1). Hence, at least some lesion nematode species are equipped with both root-knot and cyst nematode-like cellulases.

The positioning of *Ditylenchus spp.* is based only on one genomic and two CDSs sequences, and should be considered as a consequence of the virtual absence of cellulases data from other, more basal Tylenchida. The current positioning is not well supported, and will not be discussed.

Exon-intron structure

The number of introns in the catalytic domains of the nematode cellulases varied between zero and seven. For the intron identifiers, we followed the nomenclature proposed by Kyndt et al. (2008) [20]. As a consequence of the increase in the number and the diversity of catalytic domains additional predicted intron positions were found. To label new intron positions without uprooting the existing systematics, we used identifiers such as 1½ and 5½ for introns positioned between introns 1 and 2, introns 5 and 6, *etc.* (see Additional file 2: Table S2).

Most of the *in silico* predicted introns in the newly generated cellulase sequences were located at positions equivalent to the positions reported before for plant parasitic cyst and root-knot nematodes ([28], [15], [26] and [20]) and the outgroup *Aphelenchus avenae* [29]. Identifiers 1½, 12½, 15½ were added as - up to now - unique introns in cellulase gDNAs from *A. avenae*, whereas 4½ and 5½ were used to indicate new intron positions in *Hl-eng1* and *Pv-eng2*. Among the newly generated cellulase sequences the largest intron was found in *Gp-eng1*; intron 2 with a length of 563 bp. Numerous occasions of intron gain and loss were observed in all three main types of cellulase catalytic domains tree (A, B and C). Particular introns appear to be characteristic for catalytic domain types: Type A typically contains introns 4 and 14, whereas all type B and C catalytic domains investigated so far have lost intron 18. Type C representatives share the presence of intron 2. However, this feature is not unique as it is typical for the type B cellulases from cyst nematodes as well. Among root-knot nematode type B cellulase the presence of intron 1 appears as a common characteristic. The two other cellulase catalytic domains with an intron at position 1 were found in a type B catalytic domain *Pp-eng-5* from *Pratylenchus penetrans* and interestingly, in a type A catalytic domain *Hg-eng-6* from soybean cyst nematodes.

Intron phase distribution

The intron phase distribution in the catalytic domain of nematode GHF5 cellulases was biased towards phase 0; 16 out of the 24 intron positions (66%) were inserted in between two codons. Respectively, two and four phase 1 (after the first base of a codon) and phase 2 (after the second base of a codon) introns were identified, whereas intron positions 7 and 17 occurred in two phases (0 and 1). To some extent a bias towards the phase 0 introns was to be expected as the overall frequencies of intron phases 0, 1 and 2 in *Caenorhabditis elegans* are roughly 50%, 25% and 25% [30]. This phase bias seems to be stronger in the case of cellulase catalytic domain from Tylenchida.

In case of mixed phase intron 7, phase 1 was observed only for *Aphelenchus avenae*, a fungivorous nematode that can feed on root hair or epidermal cells of plants as well. Contrary to all other taxa investigated here, *A. avenae* does not belong to the order Tylenchida, though it is included in the infraorder Tylenchomorpha ([31]). A phase 0 variant of intron 7 is found among a subset of the type A catalytic domains: the ones present in *Meloidogyne incognita* (with one exception: *Minc 19090a*), and in *Pratylenchus vulnus*. Among the 47 taxa harboring mixed phase intron 17, 13 were in phase 1, and 34 in phase 0. Phase 1 appeared to be a typical characteristic for type B and type C catalytic domains of cyst nematodes. Hence, it was also present in *eng3* and *eng4* from *G. rostochiensis*, two otherwise highly distinct cellulases. The only other case of an intron 17 in phase 1 was observed for *eng-1B* from the banana root nematode *Radopholus similis* (family Pratylenchidae). This could be seen as a

confirmation of a recent SSU rDNA-based analysis suggesting a (unexpected) close relatedness between *Radopholus* and cyst nematodes [18].

Conclusions

Addition of 45 new genomic sequences from the catalytic domain of cellulases from plant parasitic members of the order Tylenchida, followed by phylogenetic analyses further develops our understanding of the evolution of cellulases within a nematode order that harbors most economically high impact plant parasites. Three distinct types of catalytic domains were distinguished, and we hypothesize that types of catalytic domains reflect distinct substrate preferences. Numerous plant parasitic nematode species were shown to harbor two types of GHF5 cellulases. *Heterodera glycines*, soybean cyst nematode, is the only example of a plant parasite equipped with all three types of catalytic domains distinguished so far.

All Clade 12 members of the phylum Nematoda analyzed to date harbor one or multiple GHF5 cellulases. This also holds for basal representatives that are not fully dependent on plants as sole food source. *Aphelenchus spp.*, on the basis of SSU rDNA data suggested to be sister to all Tylenchida [21], is primarily mycetophagous (fungal cell walls do not contain cellulose), but can also grow and multiply on various plant species [32]. It is noted that the necromenic nematode species *Pristionchus pacificus* (Clade 9, Diplogasteridae) is harbouring seven cellulase genes all belonging to GHF5 ([33]). Hence, although GHF5 cellulases are widespread within Clade 12, they are not exclusively present in this clade.

The family Anguinidae harbours the most ancestral representatives of the order Tylenchida included in this paper, and *Ditylenchus destructor*, a member of this family and the causal agent of dry rot in potato tuber, is also known to feed on fungal hyphae. Recently, another early branching representative of the Tylenchida, *Deladenus siridicola* - a nematode with a mycetophagous and an insect parasitic life cycle, was shown to harbour GHF5 cellulases (dr. Bernard Slippers and co-workers, pers. comm.). Hence, all members of Clade 12 seem to harbor GHF5 cellulases, even ones that according to literature do not feed on plants. Presuming GHF5 cellulases were indeed acquired by lateral gene transfer, the most parsimonious explanation of the current cellulase tree would be the acquisition of such a gene by an ancient representative of the Pratylenchidae. Though we realize that current datasets are too fragmented to make a strong statement, our results are compatible with a scenario in which a GHF5 cellulase was acquired by the common ancestor of *Aphelenchus* and all Tylenchida, followed by one or multiple gene duplications and subsequent diversification.

Competing interest

The authors have declared that there are no competing interests.

Authors' contributions

KR-M and JH designed this study, HvM collected and identified the biological material. KR-M, HRM, SvdE carried out the molecular work. KR-M and PM did the phylogenetic analysis. KR-M, GS, JB and JH contributed to data analyses and preparation of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

HRM wish to thank the Erasmus Mundus Eumaine programme for financial support. The authors wish to thank Erwin Roze for sequence information on *Meloidogyne chitwoodi* cellulases. Moreover, we wish to thank dr. Yiannis Kourmpetis and dr. Edouard Severing from Laboratory of Bioinformatics (Wageningen University, The Netherlands).

References

1. Woolfit M, Iturbe-Ormaetxe I, McGraw EA, O'Neill SL: **An ancient horizontal gene transfer between mosquito and the endosymbiotic bacterium *Wolbachia pipientis***. *Mol Biol Evol* 2009, **26**(2):367–374.
2. Smant G, Stokkermans J, Yan YT, de Boer JM, Baum TJ, Wang XH, Hussey RS, Gommers FJ, Henrissat B, Davis EL, *et al*: **Endogenous cellulases in animals: Isolation of beta-1,4-endoglucanase genes from two species of plant-parasitic cyst nematodes**. *Proc Natl Acad Sci USA* 1998, **95**(9):4906–4911.
3. Keen NT, Roberts PA: **Plant parasitic nematodes: Digesting a page from the microbe book**. *Proc Natl Acad Sci USA* 1998, **95**(9):4789–4790.
4. Popeijus H, Overmars H, Jones J, Blok V, Goverse A, Helder J, Schots A, Bakker J, Smant G: **Enzymology: Degradation of plant cell walls by a nematode**. *Nature* 2000, **406**(6791):36–37.
5. Jaubert S, Laffaire JB, Abad P, Rosso MN: **A polygalacturonase of animal origin isolated from the root- knot nematode *Meloidogyne incognita***. *FEBS Lett* 2002, **522**(1–3):109–112.
6. Mitreva-Dautova M, Roze E, Overmars H, De Graaff L, Schots A, Helder J, Goverse A, Bakker J, Smant G: **A symbiont-independent endo-1,4- β -xylanase from the plant-parasitic nematode *Meloidogyne incognita***. *Mol Plant Microbe Interact* 2006, **19**(5):521–529.
7. Qin L, Kudla U, Roze EHA, Goverse A, Popeijust H, Nieuwland J, Overmars H, Jones JT, Schots A, Smant G, *et al*: **A nematode expansin acting on plants**. *Nature* 2004, **427**(6969):30.
8. Maggenti AR: **Nemic relationships and the origin of plant parasitic nematodes**. In *Plant parasitic nematodes*. Edited by Zuckerman BM, Mai WF, Rohde RA. New York: Academic Press Inc; 1971:65–81.
9. Ford Doolittle W: **You are what you eat: A gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes**. *Trends Genet* 1998, **14**(8):307–311.
10. Gladyshev EA, Meselson M, Arkhipova IR: **Massive horizontal gene transfer in bdelloid rotifers**. *Science* 2008, **320**(5880):1210–1213.

11. Reardon W, Chakrabortee S, Pereira TC, Tyson T, Banton MC, Dolan KM, Culleton BA, Wise MJ, Burnell AM, Tunnacliffe A: **Expression profiling and cross-species RNA interference (RNAi) of desiccation-induced transcripts in the anhydrobiotic nematode *Aphelenchus avenae*.** *BMC Mol Biol* 2010, **11**:6.
12. Wharton DA, Aalders O: **Desiccation stress and recovery in the anhydrobiotic nematode *Ditylenchus dipsaci* (Nematoda: Anguinidae).** *Eur J Entomol* 1999, **96**(2):199–203.
13. Shannon AJ, Tyson T, Dix I, Boyd J, Burnell AM: **Systemic RNAi mediated gene silencing in the anhydrobiotic nematode *Panagrolaimus superbis*.** *BMC Mol Biol* 2008, **9**:58.
14. Abad P, Gouzy J, Aury JM, Castagnone-Sereno P, Danchin EGJ, Deleury E, Perfus-Barbeoch L, Anthouard V, Artiguenave F, Blok VC, *et al*: **Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*.** *Nat Biotechnol* 2008, **26**(8):909–915.
15. Danchin EGJ, Rosso MN, Vieira P, De Almeida-Engler J, Coutinho PM, Henrissat B, Abad P: **Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes.** *Proc Natl Acad Sci USA* 2010, **107**(41):17651–17656.
16. Opperman CH, Bird DM, Williamson VM, Rokhsar DS, Burke M, Cohn J, Cromer J, Diener S, Gajan J, Graham S, *et al*: **Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism.** *Proc Natl Acad Sci USA* 2008, **105**(39):14802–14807.
17. Bert W, Leliaert F, Vierstraete AR, Vanfleteren JR, Borgonie G: **Molecular phylogeny of the Tylenchina and evolution of the female gonoduct (Nematoda: Rhabditida).** *Mol Phylogenet Evol* 2008, **48**(2):728–744.
18. Holterman M, Karssen G, van den Elsen S, van Megen H, Bakker J, Helder J: **Small subunit rDNA-based phylogeny of the Tylenchida sheds light on relationships among some high-impact plant-parasitic nematodes & the evolution of plant feeding.** *Phytopathology* 2009, **99**(3):227–235.
19. Tandingan De Ley I, De Ley P, Vierstraete A, Karssen G, Moens M, Vanfleteren J: **Phylogenetic analyses of *Meloidogyne* small subunit rDNA.** *J Nematol* 2002, **34**(4):319–327.
20. Kyndt T, Haegeman A, Gheysen G: **Evolution of GHF5 endoglucanase gene structure in plant-parasitic nematodes: No evidence for an early domain shuffling event.** *BMC Evol Biol* 2008, **8**:305.
21. Van Megen H, Van Den Elsen S, Holterman M, Karssen G, Mooyman P, Bongers T, Holovachov O, Bakker J, Helder J: **A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences.** *Nematology* 2009, **11**(6):927–950.
22. Holterman M, van der Wurff A, van den Elsen S, van Megen H, Bongers T, Holovachov O, Bakker J, Helder J: **Phylum-wide analysis of SSU rDNA reveals deep phylogenetic**

relationships among nematodes and accelerated evolution toward crown clades. *Mol Biol Evol* 2006, **23**(9):1792–1800.

23. Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucleic Acids Symp Ser* 1999, **41**:95–98.

24. Rambaut A, Drummond A: *Tracer v1.4*. 14th edition.; . <http://beast.bio.ed.ac.uk/Tracer> 2007.

25. Stamatakis A, Hoover P, Rougemont J: **A rapid bootstrap algorithm for the RAxML web servers.** *Syst Biol* 2008, **57**(5):758–771.

26. Gao B, Allen R, Davis EL, Baum TJ, Hussey RS: **Developmental expression and biochemical properties of a β -1,4-endoglucanase family in the soybean cyst nematode, *Heterodera glycines*.** *Mol Plant Pathol* 2004, **5**(2):93–104.

27. Feng Y, Duan CJ, Pang H, Mo XC, Wu CF, Yu Y, Hu YL, Wei J, Tang JL, Feng JX: **Cloning and identification of novel cellulase genes from uncultured microorganisms in rabbit cecum and characterization of the expressed cellulases.** *Appl Microbiol Biotechnol* 2007, **75**(2):319–328.

28. Yan YT, Smant G, Stokkermans J, Qin L, Helder J, Baum T, Schots A, Davis E: **Genomic organization of four beta-1,4-endoglucanase genes in plant-parasitic cyst nematodes and its evolutionary implications.** *Gene* 1998, **220**(1–2):61–70.

29. Karim N, Jones JT, Okada H, Kikuchi T: **Analysis of expressed sequence tags and identification of genes encoding cell-wall-degrading enzymes from the fungivorous nematode *Aphelenchus avenae*.** *BMC Genomics* 2009, **10**:525.

30. Coghlan A, Wolfe KH: **Origins of recently gained introns in *Caenorhabditis*.** *Proc Natl Acad Sci USA* 2004, **101**(31):11362–11367.

31. De Ley P, Blaxter ML: **Systematic Position and Phylogeny.** In *The Biology of Nematodes*. Edited by Lee DL. London, UK: Taylor & Francis; 2002:1–30.

32. Hunt DJ: *Aphelenchida, Longidoridae and Trichodoridae: their systematics and bionomics*. Wallingford, UK: CAB International; 1993.

33. Schuster LN, Sommer RJ: **Expressional and functional variation of horizontally acquired cellulases in the nematode *Pristionchus pacificus*.** *Gene* 2012, **506**(2):274–282.

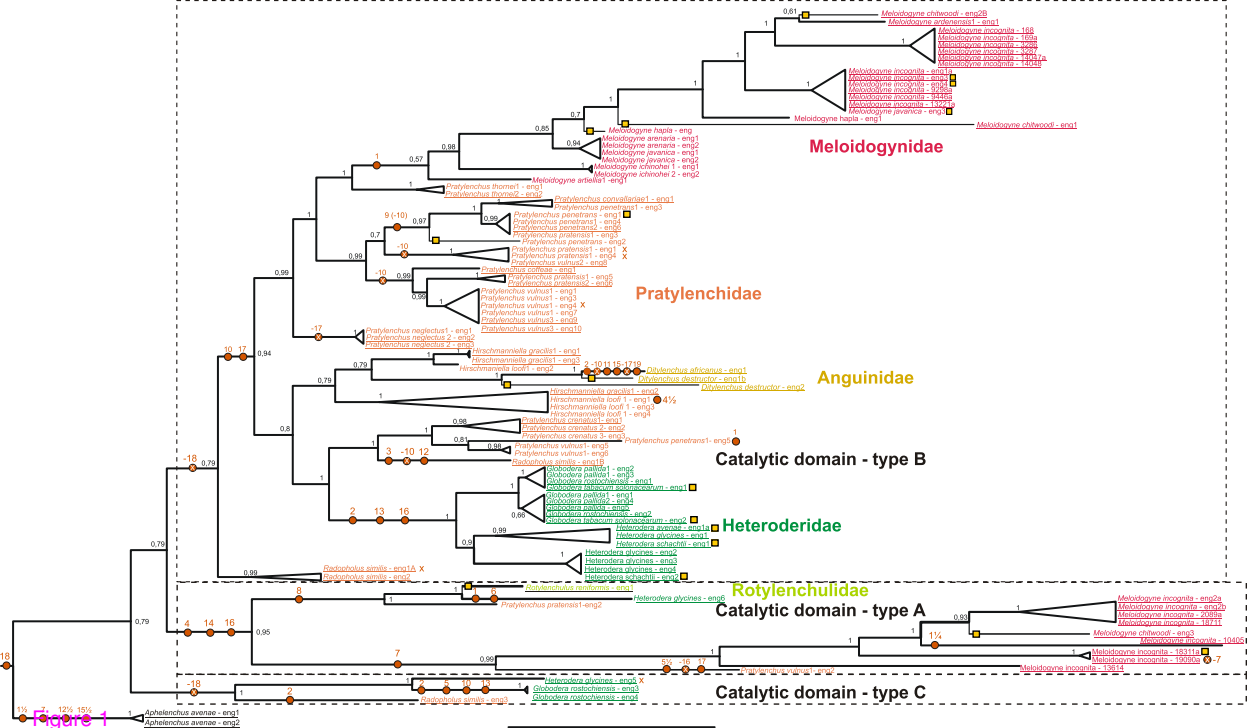
Additional files

Additional_file_1 as PDF

Additional file 1 Table S1: List of GHF5 endoglucanase sequences from plant parasitic nematodes from public databases used in this paper

Additional_file_2 as PDF

Additional file 2 Table S2: Schematic overview of the (predicted) introns in genomic sequences from the GHF5 endoglucanase genes. For the intron identifiers, we adhered to the nomenclature proposed by Kyndt et al. (2008) [20]. Identifiers such as 1½ and 5½ were used for novel introns positioned between introns 1 and 2, 5 and 6, etc. The colour scheme for the phase of the introns is explained below this Table. Length of the introns (in bp) is given inside each box



Additional files provided with this submission:

Additional file 1: 3898072797623212_add1.pdf, 93K

<http://www.biomedcentral.com/imedia/1571208455853735/supp1.pdf>

Additional file 2: 3898072797623212_add2.pdf, 97K

<http://www.biomedcentral.com/imedia/1793367373853735/supp2.pdf>