



ELSEVIER

Gene 220 (1998) 61–70

GENEAN INTERNATIONAL JOURNAL ON
GENES AND GENOMES

Genomic organization of four β -1,4-endoglucanase genes in plant-parasitic cyst nematodes and its evolutionary implications

Yitang Yan ^a, Geert Smant ^{b,1}, Jack Stokkermans ^b, Ling Qin ^b, Johannes Helder ^b,
Thomas Baum ^c, Arjen Schots ^b, Eric Davis ^{a,*}

^a Plant Pathology Department, North Carolina State University, Raleigh, NC 27695, USA

^b Nematology, Wageningen Agricultural University, Binnenhaven 10, 6709 PD Wageningen, The Netherlands

^c Plant Pathology Department, Iowa State University, Ames, IA 50011, USA

Received 6 July 1998; accepted 31 July 1998; Received by A. Nakazawa

Abstract

The genomic organization of genes encoding β -1,4-endoglucanases (cellulases) from the plant-parasitic cyst nematodes *Heterodera glycines* and *Globodera rostochiensis* (HG-*eng1*, Hg-*eng2*, GR-*eng1*, and GR-*eng2*) was investigated. HG-*eng1* and GR-*eng1* both contained eight introns and structural domains of 2151 and 2492 bp, respectively. HG-*eng2* and GR-*eng2* both contained seven introns and structural domains of 2324 and 2388 bp, respectively. No significant similarity in intron sequence or size was observed between HG-*eng1* and HG-*eng2*, whereas the opposite was true between GR-*eng1* and GR-*eng2*. Intron positions among all four cyst nematode cellulase genes were conserved identically in relation to the predicted amino acid sequence. HG-*eng1*, GR-*eng1*, and GR-*eng2* had several introns demarcated by 5'-GC...AG-3' in the splice sites, and all four nematode cellulase genes had the polyadenylation and cleavage signal sequence 5'-GAUAAA-3'—both rare occurrences in eukaryotic genes. The 5'-flanking regions of each nematode cellulase gene, however, had signature sequences typical of eukaryotic promoter regions, including a TATA box, bHLH-type binding sites, and putative silencer, repressor, and enhancer elements. Database searches and subsequent phylogenetic comparison of the catalytic domain of the nematode cellulases placed the nematode genes in one group, with Family 5, subfamily 2, glycosyl hydrolases from Scotobacteria and Bacillaceae as the most homologous groups. The overall amino acid sequence identity among the four nematode cellulases was from 71 to 83%, and the amino acid sequence identity to bacterial Family 5 cellulases ranged from 33 to 44%. The eukaryotic organization of the four cyst nematode cellulases suggests that they share a common ancestor, and their strong homology to prokaryotic glycosyl hydrolases may be indicative of an ancient horizontal gene transfer. © 1998 Published by Elsevier Science B.V. All rights reserved.

Keywords: Animal cellulase; Gland secretions; Gene transfer; Intron/exon sequence; Phylogeny; Transcriptional regulation

1. Introduction

The plant cell wall is a primary barrier in the local environment of plant pathogens. Digestion of plant cell walls by secreted microbial exo- and endoglucanases

(cellulases) may enhance direct penetration of plant tissue, facilitate the degradation of associated cell wall components by other enzymes, and enhance the release of plant cell constituents that can be utilized for nutrition of phytopathogens (Barras et al., 1994; Walton, 1994). Cellulases and xylanases have been grouped into a number of distinct families based upon amino acid sequence identities and hydrophobic cluster analysis (Gilkes et al., 1991; Henrissat and Bairoch, 1993). Typically, these glycanases consist of a catalytic domain and a cellulose-binding domain (CBD) that is connected by a Pro/Thr/Ser-rich peptide linker (Gilkes et al., 1991). The absence of a CBD and associated peptide linker has been observed in some endoglucanases, and this appears to affect substrate specificity but does not preclude enzyme activity (Gilkes et al., 1991; Wang and Jones, 1995a,b).

* Corresponding author. Tel: +1 919 515 6692; Fax: +1 919 515 7716; e-mail: eric_davis@ncsu.edu

¹ The first two authors contributed equally to this research.

Abbreviations: aa, amino acid(s); bp, base pairs(s); CBD, cellulose-binding domain; cDNA, DNA complementary to RNA; GR-*eng*, *Globodera rostochiensis* β -1,4-endoglucanase genes; HG-*eng*, *Heterodera glycines* β -1,4-endoglucanase genes; kb, kilobase(s); mAb, monoclonal antibodies; nt, nucleotide; PCN, potato cyst nematode; PCR, polymerase chain reaction; RACE, rapid amplification of cDNA ends; SCN, soybean cyst nematode; TSP, transcription start point, *tsp*; UTR, untranslated regions.

The first isolation of endogenous cellulase (β -1,4-endoglucanase) genes from animals has recently been reported for two species of plant-parasitic cyst nematodes, *Heterodera glycines* and *Globodera rostochiensis* (Smant et al., 1998). It is hypothesized that these microscopic worms secrete cellulases that are produced in their esophageal glands through their stylet (feeding spear) to promote their migration through plant root cells. In both *H. glycines* and *G. rostochiensis*, two cellulase genes were identified. Each nematode species contained one larger (approximately 1.6 kb) endoglucanase cDNA (HG-*eng1* [AF006052] and GR-*eng1* [AF004523], respectively), and one smaller (approximately 1.2 kb) endoglucanase cDNA (termed HG-*eng2* [AF006053] and GR-*eng2* [AF004716], respectively). The endogenous expression of the four endoglucanase genes specifically within the subventral glands of both nematode species was confirmed using mRNA in-situ hybridization and polyclonal sera that was raised to heterologous proteins of each of the four isolated genes (Smant et al., 1998). Each of the four expressed cyst nematode endoglucanases contained a eukaryotic secretion signal peptide, a cellulase catalytic domain, and a polyadenylated 3' terminus (Smant et al., 1998). Both HG-*eng1* and GR-*eng1* cDNAs also contained a peptide linker and a CBD at the C-terminus (Smant et al., 1998). The predicted amino acid sequence homologies and hydrophobic cluster analyses revealed that the catalytic domains of the four nematode endoglucanases each contained the glycosyl hydrolase Family 5 signature found in some bacterial and fungal endoglucanases (Henrissat and Bairoch, 1993, 1996; Smant et al., 1998).

The cyst nematode endoglucanases only showed a significant amino acid homology with endoglucanases of bacteria. In order to determine homologies that extended beyond the coding sequences, we investigated the exon/intron organization, and 5'- and 3'-flanking regions of the genes. Although ample coding sequences and regulatory sequences have been identified in the free-living nematode, *Caenorhabditis elegans* (Riddle et al., 1997), the genomic organization and potential regulatory regions from plant-parasitic nematode genes have not been studied extensively. This report presents the genomic organization of HG-*eng1*, HG-*eng2*, GR-*eng1*, and GR-*eng2*, and highlights putative regulatory elements in the flanking sequences of these cyst nematode endoglucanase genes. In addition, our genomic study strongly suggests that the four nematode cellulases share a common ancestor, and that this putative ancestor may have acquired the cellulases from a prokaryote via horizontal gene transfer.

2. Materials and methods

Preparasitic second-stage juveniles (J2) of *Heterodera glycines* and *Globodera rostochiensis* were hatched from

eggs that were extracted from greenhouse cultures as described previously (Goverse et al., 1994; De Boer et al., 1996). Total DNA was extracted from J2 by alkaline/SDS lysis and phenol/chloroform extraction (Sulston and Hodgkin, 1988).

2.1. *H. glycines* genomic library construction and screening

The *H. glycines* genomic DNA was partially digested with *Sau3A*, *BglII* and *BamHI* and used to produce a genomic library by cloning into the Lambda FIX II/*XhoI* partial fill-in vector (Stratagene, La Jolla, CA). The *H. glycines* genomic library consisted of 76 000 primary recombinants and had an amplified titre of 430 000 pfu/ μ l. Approximately 10 000 pfu of the amplified library in 600 μ l of XL1-Blue MRA-P2 host cells were plated on 150-mm LB plates and incubated at 37°C for 8 h. Plaque lifts onto Hybond-N membranes (Amersham International, Buckinghamshire, UK) and subsequent treatments were conducted as suggested for the Genius[™] System (Boehringer Mannheim Biochemicals, Indianapolis, IN). Probes used for library screening consisted of full-length cDNA of either HG-*eng1* or HG-*eng2* cDNA that were labeled with digoxigenin (DIG) using a random-prime DIG DNA labeling and Detection Kit (Boehringer Mannheim Biochemicals). Probes were hybridized to plaque lifts at 59°C with subsequent washing at 68°C. Positive plaques were detected by binding of alkaline phosphatase-labeled anti-DIG antibody and subsequent color reaction with NBT/BCIP substrate.

2.2. Lambda DNA purification and HG-*eng1* and HG-*eng2* subcloning

Lambda Fix II DNA harboring positive genomic HG-*eng1* and HG-*eng2* clones were amplified in *E. coli* host XL1-Blue MRA-P2, lysed, and purified using a Lambda Mini Kit (Qiagen, Valencia, CA). Inserts of *H. glycines* genomic DNA were excised from the purified lambda DNA with *NotI*, and the inserts were digested with various restriction endonucleases for agarose gel electrophoresis and Southern blots (Sambrook et al., 1989). The DIG-labeled HG-*eng1* and HG-*eng2* cDNA probes were used in these Southern blots to identify restriction enzymes that yielded positive *H. glycines* genomic DNA fragments for subcloning. Genomic fragments containing HG-*eng1* or HG-*eng2* genes were subcloned into pBluescript-KS⁺, and the orientation of each clone was determined using T3 and T7 primers that were labeled with the DIG Oligonucleotide 3'-End Labeling Kit (Boehringer Mannheim Biochemicals). Both the coding and flanking regions of HG-*eng1* and HG-*eng2* were prepared for automated nucleic acid sequencing. The 5'- and 3'-flanking regions of each gene were obtained for

sequencing by PCR using one outward primer derived from the sequence of each end of the gene and a second primer derived from the appropriate flanking vector sequence. Amplified products were cloned into TA-vectors (Invitrogen, San Diego, CA) and sequenced from both directions.

2.3. Isolation of GR-eng1 and GR-eng2 genomic clones and 5'- and 3'-flanking regions

Inward primers were designed from the most distal sequences in the GR-eng1 and GR-eng2 cDNAs to amplify the genomic sequence of each gene using *G. rostochiensis* DNA as template. In order to clone the 5'- and 3'-flanking regions of GR-eng1 and GR-eng2, a genomic library of *G. rostochiensis* J2 was constructed in the plasmid vector pZErO-2 (Invitrogen). Genomic DNA of *G. rostochiensis* was partially digested with *Sau3A* to yield fragments ranging from 500 to 4000 bp, and these were subsequently ligated to a *Bam*HI-digested pZErO-2 vector for library construction. The genomic library contained 1×10^6 primary recombinants with an average insert size of 1.0 kb. The 5'- and 3'-flanking regions of the GR-eng1 and GR-eng2 genes were obtained from library clones by the same PCR strategy as used for HG-eng1 and HG-eng2.

3. Results

3.1. Genomic clones of cellulase genes from *H. glycines* and *G. rostochiensis*

Five positive plaques were each initially obtained for HG-eng1 and HG-eng2 from the genomic library, and after various restriction digestions of these clones and hybridizations with corresponding cDNAs, positive lambda Fix II subclones of each gene were selected for further analysis (Fig. 1). The location and orientation of the SCN cellulase genes in the lambda DNA were verified by PCR using cellulase-specific internal primers and the T3 and T7 primers located in the lambda vector (data not shown). For obtaining the 5'-flanking regions and coding sequence of HG-eng1, the 3.8-kb *Bam*HI digested fragment from Lambda DNA was cloned into the pBluescript KS⁺ *Bam*HI site. The 6-kb fragment containing 3'-flanking regions of HG-eng1 was cloned into the pBluescript KS⁺ *Not*I and *Bam*HI site. A 5.5-kb DNA fragment from Lambda DNA excised by *Pst*I and *Not*I, which contained both 5'- and 3'-flanking regions of HG-eng2, was cloned into the pBluescript-KS⁺ *Pst*I and *Not*I site. The lengths of the genomic structural fragments of HG-eng1 (AF052733) and HG-eng2 (AF052734) were 2151 and 2324 bp, respectively. The two genomic structural fragments encoding GR-eng1 (AF056110) and GR-eng2 (AF056111) were 2492 bp

and 2388 bp, respectively, and they were PCR-amplified directly from *G. rostochiensis* J2 genomic DNA.

3.2. Cis splicing in cyst nematode endoglucanase genes

The intron/exon boundaries of the genomic sequences of the four cellulase genes were determined by aligning the genomic sequence with the cDNA sequence for each corresponding gene (Fig. 2). Both *eng1* genes contained eight introns, and both *eng2* genes contained seven introns. The intron size among the four genes ranged from 45 to 742 bp, with an average intron size of 128 bp. The exon size among the four genes ranged from 70 to 311 bp, with an average exon size of 158 bp. The average G+C values of the exons in GR-eng and HG-eng were 53.8 and 49.5%, respectively. The average G+C values of the introns in GR-eng and HG-eng were 42.3 and 35.5%, respectively.

The size of the corresponding introns between HG-eng1 and HG-eng2 was quite different, with HG-eng1 having smaller introns at each position. In contrast, the size of corresponding introns between GR-eng1 and GR-eng2 was highly conserved except for intron number 1. All of the intron sequences from *H. glycines* and *G. rostochiensis* endoglucanase genes were examined to identify potential homology between introns. No significant sequence homology was detected between the introns from HG-eng1 and HG-eng2, and no homology was found between HG-eng introns and GR-eng introns. However, a strong sequence homology existed between the corresponding introns of GR-eng1 and GR-eng2. Intron numbers 2–7 of GR-eng1 and GR-eng2 shared 87–96% identical internal sequence as well as a similarity in size. In contrast, the size of the first GR-eng1 intron is 308 bp as compared to that of GR-eng2, which is 742 bp. The G+C value of the first intron in GR-eng1 is significantly higher (57.8%) as compared to the other introns in GR-eng, which is explained by eight cytosine repeats (6–14 bp long). The divergence in the first GR-eng introns is not based solely on point mutations or small insertions—as seems the case for introns number 2–7. Computer analysis revealed that the most conserved region between the first GR-eng introns is a 41-bp stretch demonstrating 80.5% sequence identity. At the other extreme in intron variation, intron number 6 in both GR-eng1 and GR-eng2 shows hardly any divergence (96% sequence identity over the entire intron sequence). Out of the 51 nucleotides that are contained in intron 6, two point mutations (transitions) have occurred. The point mutations were transitions from A to G at +13 bp downstream of the 5'-splice site, and from C to T at –4 bp upstream of the 3'-splice site.

3.3. Intron splicing sequences, polyadenylation and cleavage signals

The *cis*-splicing in the cyst nematode cellulase genes largely complies to the 'GU-AG rule' (Blumenthal and

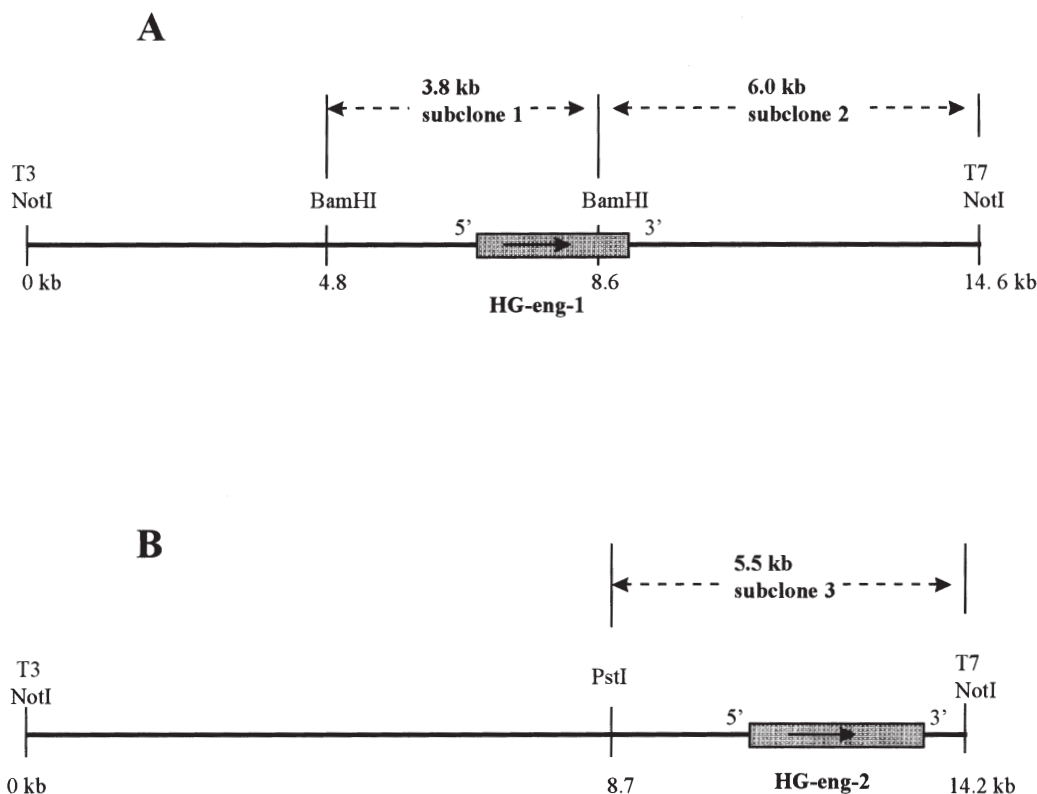


Fig. 1. Position of *Heterodera glycines* endoglucanase genes *HG-eng1* (A) and *HG-eng2* (B) on *H. glycines* genomic DNA fragments excised from lambda Fix II clones. Dark boxes with arrows represent coding regions and transcriptional orientation.

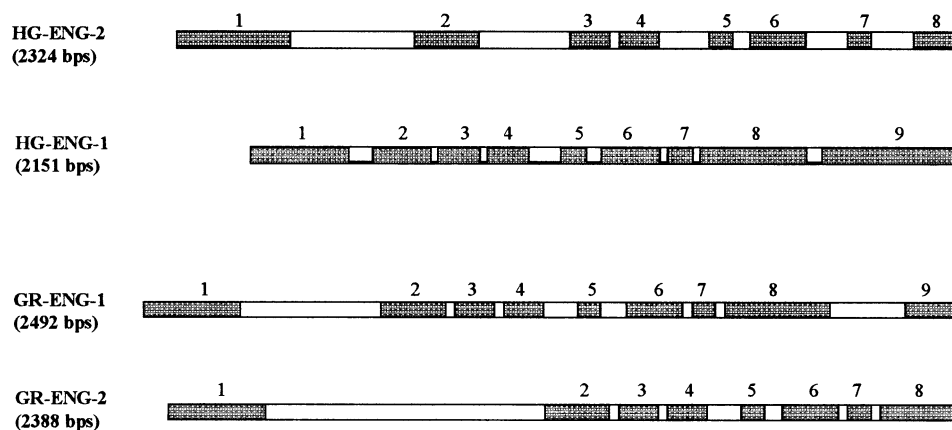


Fig. 2. Exon/intron organization of four cyst nematode endoglucanase genes, *Heterodera glycines* genes *HG-eng1* and *HG-eng2*, and *Globodera rostochiensis* genes *GR-eng1* and *GR-eng2*. Exons (dark boxes) and introns (open boxes) are drawn proportionally to indicate the relative position and size of the introns.

Steward, 1997). However, the third and fifth introns of *HG-eng1*, and the third intron of both *GR-eng1* and *GR-eng2*, use GC as a 5'-splicing donor sequence instead of GU. In *HG-eng2*, all of the exons and introns are bordered by canonical sequences. Based upon the collective exon/intron border sequences of both cellulase genes in *G. rostochiensis*, the consensus 5'-splice site is M G₇₃ | G₁₀₀U₈₇R A₆₀G₆₇U₇₃, and the consensus 3'-splice site is C₆₇C₅₃R U₇₃U₆₇A₅₃A₁₀₀G₁₀₀ | G₈₇ Y.

For both *H. glycines* cellulase genes, the consensus 5'-splice site is M G₇₃ | G₁₀₀U₈₇R A₆₀G₆₇U₇₃, and the consensus 3'-splice site is Y₇₃Y₇₃Y₇₃U₈₀U₅₃A₅₃A₁₀₀G₁₀₀ | G₈₇ U₇₃. The polyadenylation and cleavage signals for the four nematode cellulase genes are all 5'-GAUAAA-3', which is rarely used in nematodes and vertebrates (Blumenthal and Steward, 1997). The 3'-end formation signals of *HG-eng1* and *HG-eng2* are located at 32 and 17 nucleotides upstream of the cleavage site,

respectively. The 3'-end formation signals for GR-eng1 and GR-eng2 both occur at 20 nucleotides upstream of the cleavage site.

3.4. Introns are inserted at the same positions with respect to the encoded cellulases

HG-eng1 and HG-eng2 share 78.4% predicted aa identity in the catalytic domain, and the two GR-eng cellulases share 83.0% predicted aa identity in the catalytic domain (Smant et al., 1998). The two least similar cellulases of the four, HG-eng2 and GR-eng1, share 71.4% identity in aa sequence over the catalytic domain. When the cellulase protein alignment was overlaid with the exon/intron data, it was recognized that the intron positions with respect to predicted aa sequence are identical among the four nematode cellulases (Fig. 3). Cellulases from *H. glycines* had apparent amino acid deletions in positions 94 and 165, at the second and third exons, respectively.

3.5. The location of putative cis regulatory elements in nematode cellulase flanking regions

The nucleic acid sequences upstream of the transcription start point (*tsp*) were compared among the four nematode cellulase genes (Fig. 4). Comparison of the 5'-flanking regions between HG-eng1 and HG-eng2 did not reveal any conspicuous homology. In contrast, the 5'-flanking region of GR-eng1 and GR-eng2 demonstrated strong homology, with 88.5% nucleic acid identity in the 322-bp region preceding the *tsp*. When the 3'-flanking regions of the nematode cellulases were compared, no significant homologies were apparent (data not shown).

In the 5'-flanking region of all four nematode cellulase genes, a TATA box occurs at nucleotide positions ranging from -28 to -47 upstream of the *tsp* (Gannon et al., 1979). In fact, the 5'-flanking regions of the HG-eng1, HG-eng2 and GR-eng1 genes all contained two consensus TATA box sites. There were no consensus CAAT (5'-GCCAAT-3') sequences in the 5' flanking regions of the four nematode cellulase genes (Groot et al., 1991).

The DNA motifs CAAGTG and CAATTG, which were identical to the consensus core binding site sequence (CANNTG) of the bHLH (basic helix-loop-helix proteins) class of transcriptional factors, were located at positions -72 and -396 bp relative to the *tsp* of GR-eng2 (Pabo, 1992). The bHLH core sequence also occurred at the -82, -330 and -685 positions of GR-eng1. The bHLH core sequence was located at positions -277, -177, -52, and -20 of Hg-eng1. The 5'-flanking region of HG-eng2 did not contain a consensus bHLH core sequence. A putative silencer-binding-site-2 (5'-TATTA AAA-3') was present only in position

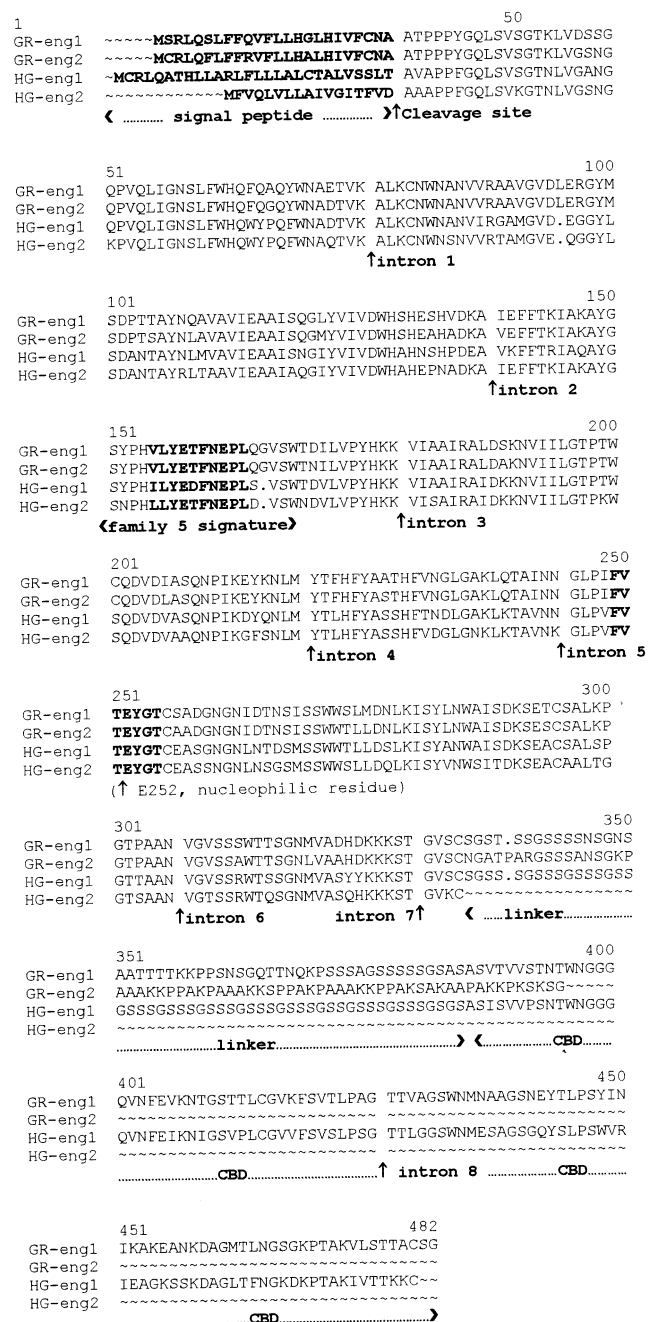


Fig. 3. Intron position with respect to the predicted amino acid sequence of the four cyst nematode endoglucanase genes, *Heterodera glycines* genes HG-eng1 and HG-eng2, and *Globodera rostochiensis* genes GR-eng1 and GR-eng2. The location of the secretion signal peptide and cleavage site, glycosyl hydrolase Family 5 signature sequence, catalytic domain (nucleophilic residue), peptide linker region, and cellulose-binding domain (CBD) of the endoglucanases are indicated.

-351 of HG-eng2 (Weissman and Singer, 1991). Putative repressor elements (consensus sequence 5'-CCWTNTTNNNW-3') were identified in positions -981 and -532 of HG-eng1 and HG-eng2, respectively. There were also putative repressor elements in the

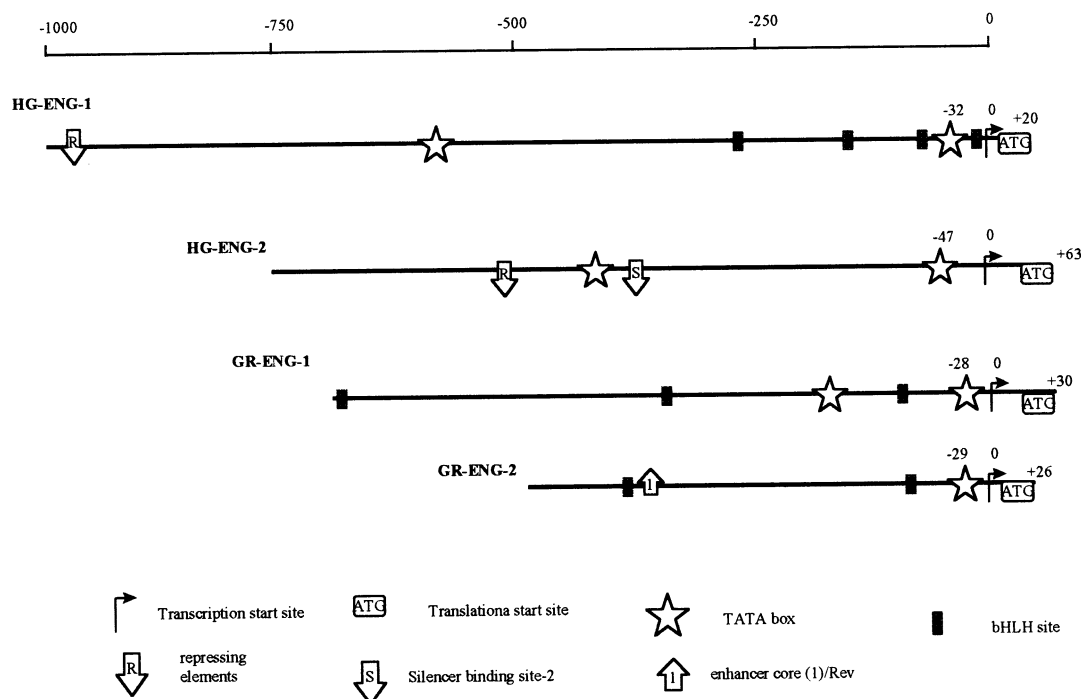


Fig. 4. Alignment of the 5'-flanking regions of the four cyst nematode endoglucanase (structural) genes, *Heterodera glycines* genes HG-*eng1* and HG-*eng2*, and *Globodera rostochiensis* genes GR-*eng1* and GR-*eng2*. Consensus transcriptional *cis* elements and their relative position with respect to the transcription start point (*tsp*) are depicted by the icons represented above.

3'-flanking sequence of GR-*eng1* at positions +49, +283, +320, and GR-*eng2* at positions +20 and +315 of the cleavage site (Shrivastava and Calame, 1994). In position -356 of GR-*eng2*, there is a putative enhancer core-1 element in reverse orientation (Weiher et al., 1983). In the 3'-flanking region of GR-*eng2*, there is a putative enhancer core-2 element (consensus sequence 5'-GGGRHTYYCC-3') located at position +55 of the polyadenylation cleavage site (Clark et al., 1988).

3.6. The relationship between cyst nematode and bacterial cellulases

The first 30 most homologous sequences in TFASTA analyses of the catalytic domain of the four cyst nematode cellulases were those of Family 5 (subfamily 2) endoglucanases of bacterial origin (Henrissat and Bairoch, 1996). The top 10 most homologous cellulases were chosen to construct a phylogenetic tree (Fig. 5). The neighbor-joining method was used to establish phylogenetic links among the various cellulases (Sudhir et al., 1993). The 14 cellulases analyzed could be divided into three related groups. The first group was the plant-parasitic cyst nematode cellulases. The second group contained CelE and celZ from *Pseudomonas* and *Erwinia*, which belong to the *Gracilicutes*, *Scotobacteria*. The third group included cellulases from the *Firmicutes*, *Bacillaceae*. The HG-*eng2* cellulase may be considered to be the most primitive type of cyst nematode cellulase

of the four analyzed. The HG-*eng2* cellulase has a catalytic domain consisting of 319 amino acids, but no peptide linker or CBD. The HG-*eng2* cellulase shares 71–78% identity in predicted amino acid sequence with the other three cyst nematode cellulase (Table 1). The HG-*eng2* cellulase shares 33–44% identity in predicted amino acid sequence with the 10 most homologous bacterial cellulases.

4. Discussion

4.1. Organization of the cyst nematode β -1,4-endoglucanase genes

We had previously reported several lines of evidence for the endogenous production of β -1,4-endoglucanases from *G. rostochiensis* and *H. glycines* (Smant et al., 1998). Herein, we showed that these four cyst nematode cellulase genes are interspersed by several introns, thus supporting the eukaryotic organization of the genes. The relatively small intron size (average intron size 128) of the four cyst nematode cellulase genes suggested that the intron size range is similar to that of *C. elegans* (Blumenthal and Steward, 1997). The average G+C value of coding and non-coding regions further points to a similarity with the genomic composition of *C. elegans*. The 5'-flanking regions of all four cyst nematode cellulase have a TATA box preceding the translation

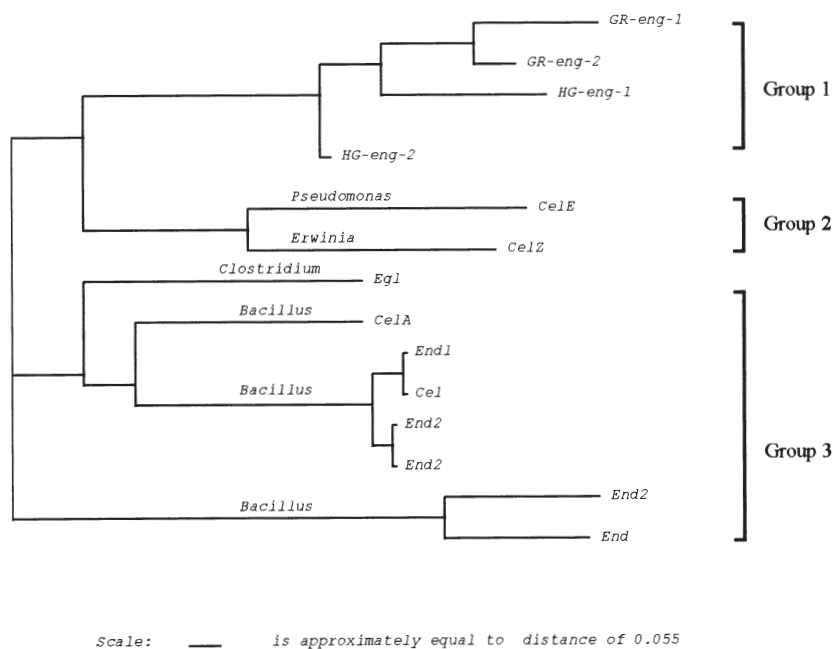


Fig. 5. Phylogenetic tree of the cyst nematode endoglucanase genes, *Heterodera glycines* genes HG-eng1 and HG-eng2, and *Globodera rostochiensis* genes GR-eng1 and GR-eng2, in relation to the most closely related groups of endoglucanases. The tree was constructed by comparisons of the catalytic domains of each endoglucanase using the neighbor-joining method of Sudhir et al. (1993).

start site in the usual location for eukaryotes (Lewin, 1994). The occasional 5'-GU...AG-3' intron splicing sites found in the cyst cellulase genes, however, are a rare occurrence in *C. elegans* and most vertebrates (Blumenthal and Steward, 1997). In addition, all four of the cyst nematode cellulase genes had the same rarely used polyadenylation and cleavage signal sequence, 5'-GATAAA-3'. By contrast, in the major sperm protein genes of *G. rostochiensis*, the polyadenylation signal was 5'-AATAAA-3' (Novitski et al., 1993). The combined features of the four cyst nematode cellulase genes suggest that whereas they are indeed similar to other nematode

genes, they may represent a unique group of eukaryotic genes.

4.2. The cis transcriptional elements of the four nematode cellulase genes

Almost no information is available on the transcriptional regulation of gene expression in plant-parasitic nematodes. The HG-eng1 and HG-eng2 cellulases were both expressed specifically within the subventral esophageal glands of infective juveniles of *H. glycines* (Smant et al., 1998), yet there was little homology between their

Table 1

Homology of the catalytic domain of the endoglucanase of *Heterodera glycines* (HG-eng-2) to the catalytic domain of other endoglucanases of cyst nematode origin, and the most closely related endoglucanases from bacteria

Enzyme	Organism	Accession No.	Homology ^a with HG-eng-2	
			Similarity (%)	Identity (%)
HG-eng-1	<i>Heterodera glycines</i>	AF006052	84.3	78.4
GR-eng-1	<i>Globodera rostochiensis</i>	AF004523	82.4	71.4
CelE	<i>Pseudomonas fluorescens</i>	x86798	63.7	44.6
CelZ	<i>Erwinia chrysanthemi</i>	y00540	61.4	41.8
Egl	<i>Clostridium acetobutylicum</i>	m31311	62.4	42.8
CelA	<i>Bacillus</i> sp. (N-4)	m14729	56.7	39.0
End1	<i>Bacillus subtilis</i> (PAP 115)	x04689	60.4	40.5
Cel	<i>Bacillus subtilis</i> (CK2)	x67044	60.4	40.5
End2	<i>Bacillus subtilis</i> (DLG)	m16185	61.0	41.9
End2	<i>Bacillus subtilis</i> (N-24)	m28332	60.6	41.3
End2	<i>Bacillus</i> sp. (KSM-635)	m27420	57.0	32.0
End	<i>Bacillus</i> sp. (KSM-64)	m84963	58.0	33.4

^aThe similarity and identity are calculated by comparing the catalytic domain amino acid sequence using the BESTFIT program in the GCG package.

putative 5' regulatory regions observed here. However, a significant homology did exist in the 5'-flanking region of the two *G. rostochiensis* cellulase genes. Computer analyses of the 5'-flanking regions of each cellulase gene revealed consensus elements of eukaryotic promoters including at least one TATA box in each gene, and the presence of bHLH-type binding sites, silencers, repressors, and enhancer elements in several of the cyst nematode cellulase genes. The existence of multiple silencer, repressor and enhancer elements in the flanking regions of the nematode cellulase genes may suggest potential regulatory roles in gland-specific gene expression and plant parasitism. Additional studies are needed to fully characterize these putative promoters from plant-parasitic nematodes and the potential *trans*-acting factors that interact with them.

4.3. Introns mark the cyst nematode cellulases at conserved locations

The intron size and sequence in any given position are quite dissimilar overall among the four cyst nematode cellulases. However, the intron-marking positions are exactly the same with regard to the encoded cellulase peptide sequence. The striking agreement of all of the intron positions among the nematode cellulases suggests that the intron positions were established before a proposed division of *H. glycines* and *G. rostochiensis* from a common ancestor. The 'intron-early' view holds that exons should represent functional or folding elements of proteins, whereas the 'intron-late' view suggests that the insertion of introns might occur at specific DNA sequences, but may not be correlated with protein sequence (Souza et al., 1996).

The reaction mechanism of glycosyl hydrolases either leads to an overall retention (e.g. Family 5) or an inversion of the anomeric configuration at the hydrolysis site. The regions around the acid catalyst (Glu.) and the nucleophile (Glu.) within nematode cellulases, as well as in Family 5 cellulases in general, are particularly conserved. Henrissat et al. (1995) have used these regions to define 'motifs' in the active site. These motifs can be found consistently in many glycosyl hydrolases that show the overall retention mechanism. Like in other cellulases, nematode cellulases have these predicted motifs—a motif near the proton donor (α -helix, loop, and β -strand; Asn–Glu), and one near the nucleophilic residue (α -helix, loop, and β -strand; Glu). Surprisingly, both motifs are not contained in a single exon each. Intron 2 is located in the middle of the α -helix of the first motif, whereas intron 5 is between the α -helix, and the β -strand of the second motif. Thus, in nematode cellulase genes the exons are not correlated to distinct modules predicted for the secondary amino acid structure. The tertiary structure predicted for the catalytic domain in the nematode cellulases is a $(\beta/\alpha)_8$ -barrel,

linked to a CBD by an 50-aa-long extended loop (Smant et al., 1998). The catalytic domains in the nematode cellulases are separated from the linker and CBD coding regions by an intron. The catalytic domain results from joining exons 1–7, whereas the linker region and the CBD are spliced from exons 8 and 9. These findings are in contrast to the cellulase gene *cel3* of the fungus *Agaricus bisporus* in which the exons are correlated not only with functional domains, but also with modular secondary structure regions of the encoded protein (Yague et al., 1996). The alignment of intron position with protein structure has previously been reported for triose-phosphate isomerase genes (encoding $(\beta/\alpha)_8$ -barrel) between organisms separated by great evolutionary distances such as maize and vertebrates (Gilbert et al., 1986). Moreover, some intron positions were found to be conserved in glycosyl hydrolase genes cloned from *Trichoderma reesi*, *Phanerochate chrysosporium*, and *A. bisporus* (Yague et al., 1996). If conservation of intron position is considered as an evolutionary event, this may provide strong evidence that the cellulase genes from *H. glycines* and *G. rostochiensis* share a common ancestral gene.

4.4. The relationship among the four cyst nematode cellulases

Evidence that HG-*eng2* is more distantly related to the other three cyst cellulases is apparent not only in the phylogenetic tree based upon the catalytic domain sequence, but also in the absence of a CBD, any rare intron donor splice site (GC instead of GU), or a 5' flanking bHLH binding sequence that were present in HG-*eng1*, GR-*eng1*, and GR-*eng2*. The two *G. rostochiensis* cellulases were the most closely related, and several lines of evidence support the hypothesis that the two GR cellulases might have been derived from recent gene duplication and exon shuffling: (1) the overall amino acid sequence identity of the catalytic domain of the GR-*eng1* and GR-*eng2* is 83%; (2) their 5'-flanking regions share 88% sequence identity in the 322-bp region preceding the putative transcription start site and; (3) the intron sequences between GR-*eng1* and GR-*eng2* tend to be highly conserved except for the first intron.

4.5. The cyst nematode cellulases and Family 5 glycosyl hydrolases

Interestingly, all of the bacterial cellulases with the highest homology to cyst nematode cellulases belong to glycosyl hydrolase Family 5, subfamily 2 (Henrissat and Bairoch, 1996). In the phylogenetic tree (Fig. 5), the nematode cellulases and Family 5 (sub 2) cellulases were classified into three closely related groups. It has been suggested that cellulases within the same families are not only structurally related but have probably evolved

from a common ancestral gene (Henrissat et al., 1995; Tomme et al., 1995). Evidence for the interspecific transfer of cellulase genes between bacteria has been reported (Guiseppi et al., 1991).

The most rigorous criterion for establishing the potential for ancient horizontal transfer is to compare a phylogenetic tree based upon a specific protein or DNA from distantly related organisms with that of the known phylogeny for the species (Syvanen, 1994). If a significant incongruency is observed between the 'protein tree' and the 'species tree', then it is possible that horizontal gene transfer has occurred. In no other nematodes, including *C. elegans*, and indeed in no other eukaryotes, was a significant homology with the cyst nematode cellulases observed at either the nucleic acid or aa level. Although functional and mechanistic convergence is common and enzyme structural convergence has probably occurred in nematodes, there are no convincing data for genuine sequence convergence of the cyst cellulases from eukaryotic genes at either the protein or DNA level (Doolittle, 1994). A strong case can be made here that the cyst nematode cellulases may have been acquired via horizontal transfer from a prokaryote, although a potential mechanism for proposed transfer is not apparent.

Acknowledgement

This research was supported by the North Carolina Agricultural Research Service, the US Department of Agriculture under NRI grant 95-37302-1918 and FAS/ICD/RSED grant 59-3148-6-003, the United Soybean Board under grant 400-41-52, and the Iowa Agriculture and Home Economics Experimental Station (J-18078; Project No. 3381). In addition, support was provided from the Dutch Potato Board, the Dutch Technology Foundation coordinated by the Life Science foundation, European Community grants BIO2-CT92-0239, BIO4-CT96-0318, and FAIR1-CT95-0905, and the North Atlantic Treaty Organization Award CRG931004.

References

- Barras, F., Gijsegem, F., Chatterjee, A.K., 1994. Extracellular enzymes and pathogenesis of soft-rot *Erwinia*. *Annu. Rev. Phytopathol.* 32, 201–234.
- Blumenthal, T., Steward, K., 1997. RNA processing and gene structure. In: Riddle, D.L., Blumenthal, T., Meyer, B.J., Priess, J.R. (Eds.), *C. elegans II*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 117–146.
- Clark, L., Pollock, R.M., Hay, R.T., 1988. Identification and purification of EBPI: a HeLa cell protein that binds to a region overlapping the 'core' of the SV40 enhancer. *Genes Dev.* 2, 991–1002.
- De Boer, J.M., Smant, G., Govere, A., Davis, E.L., Overmars, H.A., Pomp, H., Van Gent-Pelzer, M., Zilverentant, J.F., Stokkermans, J.W.P.G., Hussey, R.S., Gommers, F.J., Bakker, J., Schots, A., 1996. Secretory granule proteins from the subventral esophageal glands of the potato cyst nematode identified by monoclonal antibodies to a protein fraction from second-stage juveniles. *Mol. Plant-Microbe Interact.* 9, 39–46.
- Doolittle, R.F., 1994. Convergent evolution—the need to be explicit. *Trends Biochem. Sci.* 19, 15–18.
- Gannon, F., O'Hare, K., Perrin, F., Le Pennec, J.P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B., Chambon, P., 1979. Organization and sequences at the 5' end of cloned complete ovalbumin gene. *Nature* 278, 428–434.
- Gilbert, W., Marchionni, M., McKnight, G., 1986. On the antiquity of introns. *Cell* 46, 151–154.
- Gilkes, N.R., Henrissat, B., Kilburn, D.G., Miller, R.C., Warren, R.A.J., 1991. Domains in microbial β -1,4-glycanases: Sequence conservation, function, and enzyme families. *Microbiol. Rev.* 55, 303–315.
- Govere, A., Davis, E.L., Hussey, R.S., 1994. Monoclonal antibodies to the esophageal glands and stylet secretions of *Heterodera glycines*. *J. Nematol.* 26, 251–259.
- Groot, R.P., Karperien, M., Pals, C., Kruijer, W., 1991. Characterization of the mouse junD promoter—high basal level activity due to an octamer motif. *EMBO J.* 10, 2523–2532.
- Guiseppi, A., Aymeric, J.L., Cami, B., Barras, F., Creuzet, N., 1991. Sequence analysis of the cellulase-encoding *celY* gene of *Erwinia chrysanthemi*: a possible case of interspecies transfer. *Gene* 106, 109–114.
- Henrissat, B., Bairoch, A., 1993. New families in the classification of glycosyl hydrolases based upon amino acid sequence similarities. *Biochem. J.* 293, 781–788.
- Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.P., Davies, G., 1995. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc. Natl. Acad. Sci. USA* 92, 7090–7094.
- Henrissat, B., Bairoch, A., 1996. Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J.* 316, 695–696.
- Lewin, B., 1994. *Genes V*. Oxford University Press, Oxford, 1272 pp.
- Novitski, C.E., Brown, S., Chen, R., Corner, A.S., Atkinson, H.J., McPherson, M.J., 1993. Major sperm protein genes from *Globodera rostochiensis*. *J. Nematol.* 25, 548–554.
- Pabo, C.O., 1992. Transcription factors: Structural families and principles of DNA recognition. *Annu. Rev. Biochem.* 61, 1053–1095.
- Riddle, D.L., Blumenthal, T., Meyer, B.J., Priess, J.R. (Eds.), 1997. *C. elegans II*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sambrook, J.E., Fritsch, E.F., Maniatis, T., 1989. *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Shrivastava, A., Calame, K., 1994. An analysis of genes regulated by the multi-functional transcriptional regulator Yin Yang-1. *Nucleic Acids Res.* 22, 5151–5155.
- Smant, G., Stokkermans, J., Yan, Y., de Boer, J.M., Baum, T., Wang, X., Hussey, R.S., Davis, E.L., Gommers, F.J., Henrissat, B., Helder, J., Schots, A., Bakker, J., 1998. Endogenous cellulases in animals: cloning of expressed β -1,4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proc. Natl. Acad. Sci. USA* 95, 4906–4911.
- Souza, S., Long, M., Schoenbach, L., Roy, S.W., Gilbert, W., 1996. Intron positions correlate with module boundaries in ancient proteins. *Proc. Natl. Acad. Sci. USA* 93, 14632–14636.
- Sudhir, K., Tamura, K., Nei, M., 1993. MEGA: Molecular Evolutionary Genetics Analysis, Version 1.01. The Pennsylvania State University Press, University Park, PA.

- Sulston, J., Hodgkin, J., 1988. Methods. In: Wood, W.B. (Ed.), The Nematode *Caenorhabditis elegans*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 587–606.
- Syvanen, M., 1994. Horizontal gene transfer: evidence and possible consequences. *Annu. Rev. Genet.* 28, 237–261.
- Tomme, P., Warren, R.A.J., Gilkes, N.R., 1995. Cellulose hydrolysis by bacteria and fungi. *Adv. Micro. Physiol.* 37, 1–81.
- Walton, J.D., 1994. Deconstructing the cell wall. *Plant Physiol.* 104, 1113–1118.
- Wang, H., Jones, R.W., 1995a. Cloning, characterization and functional expression of an endoglucanase-encoding gene from the phytopathogenic fungus *Macrophomina phaseolina*. *Gene* 158, 125–128.
- Wang, H., Jones, R.W., 1995b. A unique endoglucanase-encoding gene cloned from the phytopathogenic fungus *Macrophomina phaseolina*. *Appl. Environ. Microbiol.* 61, 2004–2006.
- Weiher, H., Konig, M., Gruss, P., 1983. Multiple point mutations affecting the simian virus enhancer. *Science* 219, 626–631.
- Weissman, J., Singer, D.S., 1991. Striking similarities between the regulatory mechanisms governing yeast mating-type genes and mammalian major histocompatibility complex genes. *Mol. Cell. Biol.* 11, 4228–4234.
- Yague, E., Chow, C.M., Challen, M.P., Thurston, C.F., 1996. Correlation of exons with functional domains and folding regions in a cellulase from *Agaricus bisporus*. *Curr. Genet.* 30, 56–61.