Comp. Parasitol.
72(1), 2005, pp. 39–49

Molecular Characterization of the Actin Gene from Cyst Nematodes in Comparison with Those from Other Nematodes

ELENA S. KOVALEVA,[1,2] SERGEI A. SUBBOTIN,[3] EDWARD P. MASLER,[1] AND DAVID J. CHITWOOD[1,4]

[1] Nematology Laboratory, U.S. Department of Agriculture, Agricultural Research Service, Building 011A, Room 165B, BARC-West, Beltsville, Maryland 20705, U.S.A.,

[2] Present address: Chesapeake PERL, Inc., 387 Technology Dr., College Park, MD 20742, U.S.A., and

[3] Institute of Parasitology of Russian Academy of Sciences, Leninskii prospect 33, Moscow 117071, Russia

ABSTRACT: Actin is an abundant, highly expressed, and much conserved protein belonging to the actin–heat shock protein70–sugarkinase superfamily. The full-length messenger RNAs encoding actin were cloned and characterized from the plant-parasitic cyst nematodes *Heterodera glycines* and *Globodera rostochiensis* and from the free-living nematode *Panagrellus redivivus*. The actins from the plant-parasitic nematodes showed highest amino acid sequence identity to filarial nematode homologues, whereas nucleotide sequence identity was much lower than that for many actins from very distant organisms, such as vertebrates, plants, or fungi. Analysis of base composition revealed a striking difference between the *H. glycines* actin gene and those of filarial nematodes in (G + C) content and usage of particular codons. Analysis of genomic DNA revealed the presence of 7 introns in the *H. glycines* actin gene. The first was atypically long and started with a GC dinucleotide, and the fifth intron occupied a novel site in the catalogue of intron positions known thus far for actin genes. Analysis of the partial genomic DNA sequences obtained from 6 other *Heterodera* spp. suggested that features shown for the *H. glycines* actin gene are characteristic for a wide range of cyst nematodes.

KEY WORDS: actin, gene, (G + C) content, cyst nematode, *Heterodera glycines*, *Globodera rostochiensis*, *Panagrellus redivivus*.

Actins are highly conserved structural proteins ubiquitously expressed in eukaryotic cells. These proteins are among the most abundant in a cell, comprising as much as 1–10% of the cellular protein. The actin cytoskeleton is not only important for cell motility and chemotaxis but also constitutes a central organizer of the cell. The broad spectrum of actin-related cellular functions includes cell division, secretion, signaling, cellular shape and volume regulation, movement and phagocytosis (De Loof et al., 1996; Sutherland and Witke, 1999). Actin filaments provide strength, connections to other cells and the extracellular matrix, paths for intracellular transport, and a scaffold for generating force. Actins often function closely with enzymes and other proteins; for example, actin and heat shock protein 70 (HSP70) interact in malarial parasite–host interrelationships (Tardieux et al., 1998).

Multicellular eukaryotes contain multiple actin genes, the sequences of which are highly conserved, with pairwise percentage identities of actins in the range of 88–98%. Actin rates third among the 20 most conserved proteins present in *Caenorhabditis elegans*, *Homo sapiens*, and *Saccharomyces cerevisiae* (Copley et al., 1999), behind 2 histones. The actin sequence has been used as a valuable phylogenetic marker for a broad range of eukaryotes to corroborate results based on ribosomal DNA (rDNA) and other protein sequences (Baldauf et al., 2000; Voigt and Wostemeyer, 2000; Holland et al., 2001). More than 50 intron positions have been reported in actins of different species, and the intron pattern of the actin gene (along with intron sequences) reliably defines clades and serves as 1 of the decisive markers for phylogenetic analysis (Ohresser et al., 1997; Donnelly et al., 1999; Lee and Gye, 2001). Within the phylum Nematoda, actin genes have been described for free-living bacterial-feeding nematode species (*C. elegans* and *Plectus acuminatus*) (Files et al., 1983; Krause et al., 1989) and several filarial nematodes (*Onchocerca volvulus*, *Setaria digitata*, *Wuchereria bancrofti*, *Brugia malayi*) (Zeng and Donelson, 1992; Saverimuttu et al., 2000), but no actin sequence is known from a plant-parasitic nematode.

Cyst nematodes cause major crop destruction worldwide. Our primary focus is on the soybean cyst nematode (*Heterodera glycines*), an obligate parasite for which we have begun to characterize genes from the actin–HSP70–sugarkinase superfamily. This superfamily includes proteins that are vital both in normal growth and adaptation to environmental stress (Bork et al., 1992; Hurley, 1996); the latter is obviously of special importance for parasites.

In this study, we characterize the first actin

[4] Corresponding author. (e-mail: chitwood@ba.ars.usda.gov)

genes from *H. glycines*, the potato cyst nematode *Globodera rostochiensis*, and several other cyst nematodes.

MATERIALS AND METHODS

Nematode cultures and DNA preparation

Heterodera glycines (population NL1-RHp) and *Panagrellus redivivus* (population LKC26) were reared and collected as described by Masler et al. (1999). Second-stage juveniles (J2) of *H. glycines* and a developmentally mixed population of *P. redivivus* were used for RNA preparations. A *H. glycines* (J2 stage) complementary DNA (cDNA) library constructed in bacteriophage λ vector Uni-ZAP™ XR (Stratagene, La Jolla, California, U.S.A.) and DNA prepared from a cDNA library of preparasitic J2 of the potato cyst nematode *G. rostochiensis* Ro1 were kindly provided as described in the Acknowledgments.

Total cellular RNA was isolated from fresh or frozen tissues of both species by a standard guanidinium thiocyanate method followed by cDNA preparation with M-MLV reverse transcriptase (GIBCO BRL, Gaithersburg, Maryland, U.S.A.) for polymerase chain reaction (PCR) amplification. Alternatively, total RNA from *P. redivivus* was converted to 5′- and 3′-elongated cDNA for rapid amplification of cDNA ends (RACE), with the SMART RACE cDNA amplification kit (Clontech, Palo Alto, California, U.S.A.). Genomic DNA was extracted from 100,000 fresh J2 of *H. glycines* and from a developmentally mixed population of *P. redivivus* with the MasterPure™ DNA purification kit (Epicentre, Madison, Wisconsin, U.S.A.). Genomic DNA was also prepared from 6 other *Heterodera* species: *Heterodera schachtii* (Belgium) from the *schachtii* group, to which belongs *H. glycines*; *Heterodera latipons* (Syria) and *Heterodera avenae* (Turkey) from the *avenae* group; *Heterodera ripae* (Belgium) and *Heterodera litoralis* (New Zealand) from the *humuli* group; and *Heterodera cyperi* (Spain) from the *cyperi* group. For each species, 1 to 4 cysts were used for DNA extraction as described by Subbotin et al. (2001).

Cloning and identification of actin cDNA

The PCR and sequencing procedures have been described (Kovaleva et al., 2002). The PCR primers were randomly chosen on the basis of multiple alignment of nucleotide sequences of several known actin genes. Several positive PCR products amplified from *H. glycines* cDNA were then sequenced and overlapping sequences were assembled into a continuous sequence. The 5′- and 3′-end regions of *H. glycines* actin were amplified from a cDNA library using nested PCR with M13REV and T3 sequencing oligonucleotides as sense PCR primers for the 5′-end and M13-21 and T7 sequencing oligonucleotides as antisense PCR primers for the 3′-end, along with corresponding specific primers derived from the obtained informative sequence. Similar techniques were applied for identification of actin sequence from *P. redivivus*, but the 5′- and 3′-end regions were detected by the RACE method.

The established coding sequence (CDS) of *H. glycines* actin was used to design a pair of primers to amplify the full-length CDS as a single PCR-product. The designed pair of 5′-AATGTGCGACGAAGAAG-3′ (sense primer) and 5′-CTCAGAAGCACTTGCGG-3′ (antisense primer) was successfully used to amplify actin cDNA of *H. glycines*, *G. rostochiensis*, and *P. redivivus*. In addition, for *H. glycines* we also used a sense primer based on the 5′-noncoding region sequence located 52 nucleotides upstream from the initiator ATG. Resulting PCR products obtained from cDNA of the 3 species were cloned into a pCR 4-TOPO vector (Invitrogen, Carlsbad, California, U.S.A.), and the presence of the full CDS and the orientation of inserts were confirmed by sequencing.

Fragments of genomic DNA of *H. glycines* actin were amplified by PCR, cloned into a pCR 4-TOPO vector and sequenced. For fragments with short introns the cloning was often omitted. Four overlapping fragments spanning the entire CDS were amplified from genomic DNA of *H. glycines* to determine the contiguous sequence of the entire gene. The precise splicing sites of introns were identified by alignment of cDNA and genomic DNA sequences, and also by using freely available programs, such as GENSCANW, GENWISE, and HMMgene. Similar techniques were applied for identification of the genomic sequence from *P. redivivus*. On the basis of the established genomic sequence of actin from *H. glycines*, we designed primers to amplify a fragment about 1,000 bp long, spanning all introns but the first. Using these primers, we obtained fragments of actin genes from other species of Heteroderidae. Two to 3 clones for each species were sequenced and positions of introns were established.

Comparative analysis of nematode actins

Data on *H. glycines* were compared with all known nematode actins according to their deduced amino acid sequences, CDS, and genomic nucleotide sequences. Sequences for the free-living nematodes *C. elegans* and *P. acuminatus* and the animal-parasitic nematodes *S. digitata*, *O. volvulus*, *W. bancrofti*, and *B. malayi* were obtained from GenBank (for accession numbers see Table 1). Base composition and codon usage were estimated by the CodonW program (Pasteur Institute, Paris, France). The partial actin nucleotide sequences of nematodes (CDS only) were aligned using Clustal X 1.64, with actin sequence of *Schistosoma japonicum* (AF223400) used as the outgroup taxon. Equally weighted maximum parsimony (MP) analysis was performed using PAUP (4.0 beta version) (Swofford, 1998). Bootstrap analysis with 1,000 replicates was calculated as measures of support for individual clades for MP.

Deposition of nucleotide sequence data

Nucleotide sequence data reported in this article are available in the GenBank™ database under the accession numbers AF318603 and AY161282 (*H. glycines*), AF539593 and AY161281 (*G. rostochiensis*), AY112716 and AY380801 (*P. redivivus*), and AY443351 through AY443356 (*H. latipons*, *H. schachtii*, *H. litoralis*, *H. avenae* [Turkey], *H. ripae*, and *H. cyperi*, respectively).

RESULTS

Characterization of actin cDNA from *Heterodera glycines*

Sequence analysis of the coding region demonstrated a single open reading frame (ORF) of 1,131 nucleotides, which encoded a putative 376 amino acid protein with molecular mass of 41.8 kDa, an

Table 1. Amino acid (aa) and nucleotide (nt) homology of actin from *Heterodera glycines* and other nematodes.

Accession number	Species	aa score*	% identity	nt score*	% identity
AF318603	*Heterodera glycines*†	783	100	2,236	100
AF539593	*Globodera rostochiensis* 1†	783	100	1,782	94
AY161281	*G. rostochiensis* 2†	780	99	1,774	94
AY112716	*Panagrellus redivivus*†	772	98	1,086	87
P30163	*Onchocerca volvulus* actin 2	775	98	170	78
P30162	*O. volvulus* action 1	772	98	170	78
AAD13153	*Setaria digitata*	771	98	170	78
P10986	*Caenorhabditis elegans* actin 4	771	97	944	85
P10983	*C. elegans* actin 1/3	770	97	912	85
P10984	*C. elegans* actin 2	770	97	793	83
CAA10111	*Plectus acuminatus*	767	96	1,102	87
AAF25819	*Wuchereria bancrofti*	764	96	57	75
P90689	*Brugia malayi*	757	95	117	77

* Homology score for BLAST search.

82-nucleotide long 5′-noncoding region, and a 160-nucleotide long 3′-noncoding region with a polyA located 131 nucleotides downstream from the stop codon. Analysis of the deduced amino acid sequence revealed the presence of 3 signatures typical of a wide range of known actins. They are in positions 54–64 and 357–365 for actins signature 1 and 2, respectively, and in position 105–117 for actin and actin-related proteins signature (Fig. 1). According to the SMART program, 2 actin-like adenosine triphosphatase domains are located from E-5 to G-147 and from R-148 to R-373; a local sequence "fingerprint," which may be diagnostic of the adenine nucleotide beta-phosphate–binding pocket, was found in positions from L-9 to K-19. On the basis of programs such as PSORT II and TargetP, which predict the intracellular location of proteins by known sequence fingerprints, the discovered actin appears to be the cytoplasmic type.

Comparison of amino acid and nucleotide sequences of actins from *Heterodera glycines* and other nematodes

Complete CDSs for actins from *P. redivivus* and *G. rostochiensis* were cloned and characterized. Available full-length sequences for other nematodes were taken from the GenBank database. A multiple alignment of amino acid sequences for nematode actins is shown in Figure 1. All nematode actin sequences are 376 residues long, 348 of which are identical. The amino acid sequences for actins from *H. glycines* and actin 1 from *G. rostochiensis* appear to be 100% identical, whereas the actin 2 from *G. rostochiensis* differs from these 2 by 132-T and 153-A residues, which are not found in any other nematode actins. The nucleotide sequences for the 2 actins from *G. rostochiensis* are distinguished by 6 nucleotides, 2 of which are in the silent position. Interestingly, 2 leucine residues in positions L-95 and L-172 were encoded by ttg triplets in actin 1 and ctg triplets in actin 2.

The amino acid sequences (208–294 residues long) for actins from *H. schachtii*, *H. latipons*, *H. ripae*, and *H. cyperi* were identical to *H. glycines* actin. The actin from *H. avenae* had R in place of K-214 and the actin from *H. litoralis* had T in place of I-176 present in all other nematode actins.

Comparison of the actin sequence from *H. glycines* with full-size actins from other species revealed high amino acid similarity, ranging from 98% identity for *O. volvulus* (actin 1 and 2), *S. digitata*, and *P. redivivus* to 95% identity for *B. malayi* (Table 1). The actual number of amino acid residues different from *H. glycines* actin varies from 6 to 16 accordingly. Four residues (L-152, V-202, I-268, and M-318) are unique for all tested Heteroderinae. Meanwhile, I/V, T, L and I/V occur in respective positions in other nematode actins (Fig. 1, shaded).

Nucleotide similarity between *H. glycines* and *G. rostochiensis* was high (94% identity); the similarity for partial sequences from other *Heterodera* spp. (625–882 bp long) varied from 95% to 99%. Similarity to filarial nematodes (75–78%) was lower than to the free-living nematodes (83–87%), although the latter had lower amino acid sequence homology (Table 1). Analysis of amino acid alignment did not reveal sufficient phylogenetic signals to construct a meaningful phylogeny, unlike phylogenetic relationships between nematodes based on actin nucleotide sequences (Fig. 2).

Base composition and codon usage

The contradiction between amino acid and nucleotide homologies motivated us to analyze the base

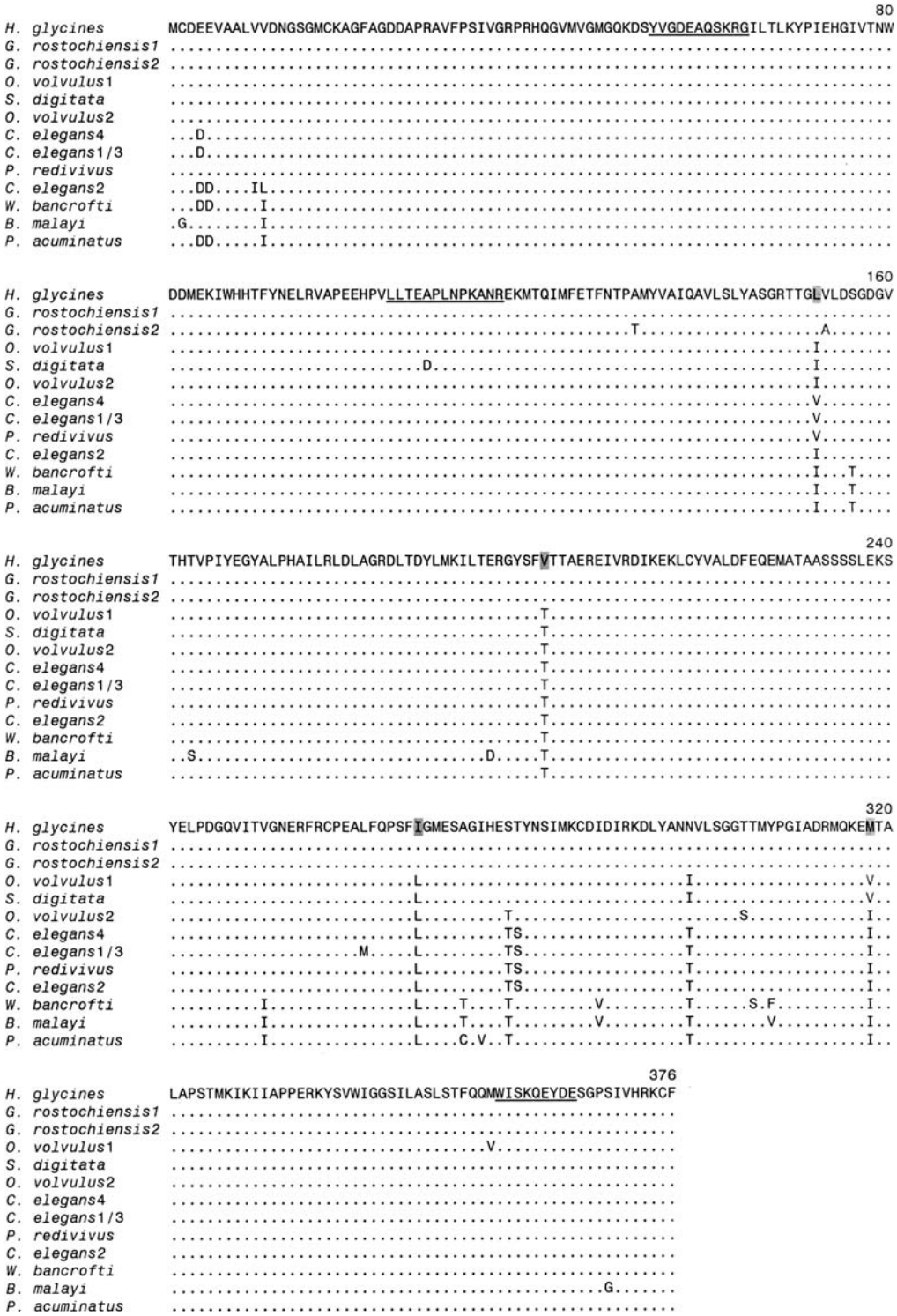

Figure 1. Alignment of the deduced amino acid sequences of actins from *Heterodera glycines* and other nematodes. Amino acid matches to the derived consensus sequence are represented by dots (.). Actins signature 1, actins and actin-related proteins signature, and actins signature 2 are underlined; the residues diagnostic for plant nematodes are shaded.

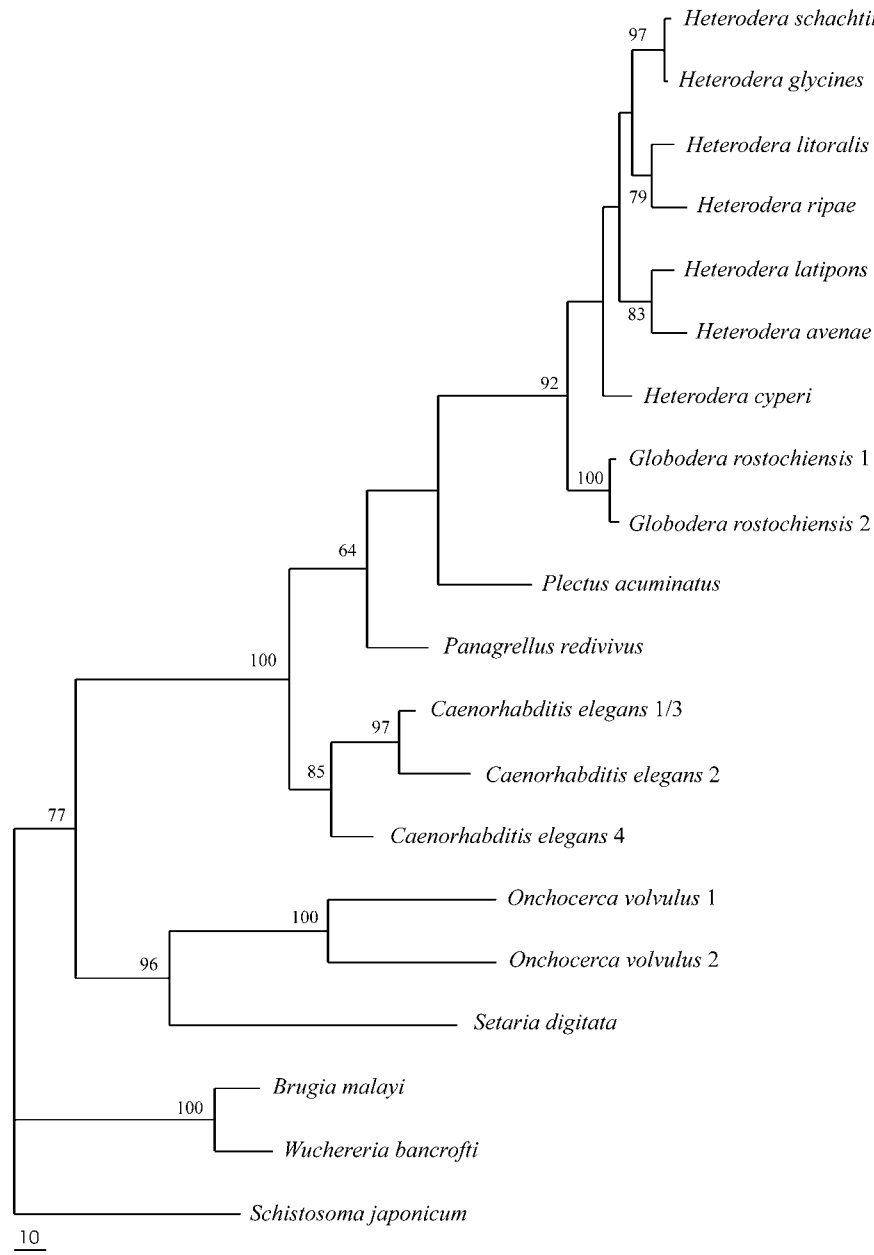

Figure 2. Phylogenetic relationships between nematodes as inferred from partial nucleotide sequences of actin gene using maximum parsimony. Bootstrap supports more than 50% are given in appropriate clades.

composition and codon usage for actin from *H. glycines* and other nematodes. The deviation for amino acid frequencies in actins from *H. glycines*, *O. volvulus*, and *S. digitata* was not greater than 1. Nonetheless, the patterns of codon usage had striking differences. The *H. glycines* actin gene preferentially used codons containing G or C at the third base position (GC3s = 0.854), whereas those of *O. volvulus* and *S. digitata* used codons ending in A or T (GC3s < 0.4). The difference in GC/AT usage at the third position between *H. glycines* and filarial nematodes was greater than 50% for 10 amino acids: Ile, Val, Ser, Pro, Thr, Ala, His, Asp, Glu, and Gly. Four amino acids (Leu, Ser, Tyr, and Cys) in *H. glycines* actin were

Table 2. The relative synonymous condon usage (RCSU) for several amino acids in actins from *Heterodera glycines* and 2 filarial nematodes.

Amino acid	Codons	*Heterodera glycines*	*Onchocerca volvulus*	*Setaria digitata*
Valine	GUU	0.17	1.39	0.83
	GUC	1.50	0.70	0.67
	GUA	0.00	1.57*	1.50*
	GUG	2.33*	0.35	1.00
Threonine	ACU	0.52	1.33	1.00
	ACC	2.61*	0.83	0.67
	ACA	0.17	1.50*	1.83*
	ACG	0.70	0.33	0.50
Alanine	GCU	0.53	2.00*	2.07*
	GCC	2.53*	0.53	0.41
	GCA	0.13	1.33	1.38
	GCG	0.80	0.13	0.14
Serine	UCU	0.00	1.62	1.38
	UCC	4.38*	1.85*	1.85*
	UCA	0.00	1.15	1.85*
	UCG	1.15	0.92	0.46
	AGU	0.00	0.23	0.23
	AGC	0.46	0.23	0.23
Leucine	UUA	0.00	0.64	0.43
	UUG	0.86	2.36*	1.93*
	CUU	0.00	0.43	1.07
	CUC	1.71	0.86	0.64
	CUA	0.00	0.43	0.86
	CUG	3.43*	1.29	1.07

100% encoded by codons ending with G or C. The relative synonymous codon usage (RSCU) values, calculated as observed frequency of a codon divided by the frequency expected for random usage, are shown for several amino acids in Table 2. Interestingly, maximum RSCU values for filarial nematodes often correspond to minimum RSCU values for *H. glycines* and vice versa.

In agreement with the very high nucleotide homology, *G. rostochiensis* and *H. glycines* have identical molar ratios of guanine + cytosine (GC content) and very similar parameters of codon preferences. The same is true for partial coding sequences from other *Heterodera* spp.

Location of introns

Comparative analysis of genomic and cDNA sequence data (Fig. 3) showed that a 1,131-bp ORF was interrupted by 7 introns in positions 19/3, 95/2, 119/3, 159/1, 226/3, 264/3, and 356/2. Corresponding analysis of genomic and cDNA sequence data for *P. redivivus* actin demonstrated the presence of 2 short introns in position 159/1 and 264/3. All introns in the *H. glycines* actin gene, except for the first one, are short (50–125 bp), concur with the major GT-AT rule, and interestingly are not AT rich (AT from 39.7 to 58.5%). The first intron is atypically long (490–493 bp) and moreover has an unusual 5′-GC instead of the standard eukaryotic GT-splicing site. The comparison of 17 individual clones spanning the first intron depicted slight inconsistency in 2 distinct locations (Fig. 3, shaded areas). The number of repetitive bases (Cs and Gs) within individual sequences varied from 8 to 11 Cs and from 10 to 14 Gs, forming several variants, the most frequent of which is shown in Figure 3.

Fragments of genomic DNA from other *Heterodera* spp. excluded the beginning of the molecule with the long first intron and corresponded to the area from the second exon to the last exon of the *H. glycines* actin gene. The samples from all *Heterodera* spp. comprised introns in identical positions to those of *H. glycines* and of similar length (from 46 to 112 bp), whereas exact sequences of those introns varied dramatically; for example, in the third intron sequence, identity to *H. glycines* was 92% for *H. schachtii* and 16% for *H. avenae*.

Comparison of intron locations in actin genes was performed for free-living and animal-parasitic nematodes for which information regarding genomic sequences was available (Fig. 4). The position of the first intron of the *H. glycines* actin gene coincides with the position of the first intron of the *C. elegans* Act4 gene. The position of the sixth intron in *H. glycines* is identical to positions of the second intron in *P. redivivus*, the third intron in *S. digitata* and both actin genes in *O. volvulus*, and is close to the position of the third introns in *B. malayi, W. bancrofti*, and *C. elegans* Act4. The seventh intron of the *H. glycines* actin gene is located identically to the last introns in *O. volvulus* and *S. digitata*.

DISCUSSION

The described actin genes from *H. glycines* and *G. rostochiensis* are the first members of the actin–HSP70–sugarkinase superfamily reported for

→

Figure 3. Genomic sequence of actin from *Heterodera glycines*. Upper-case triplets and single letters represent codon nucleotides and corresponding amino acid residues; lower-case letters represent nucleotides of introns. The 5′-GC splice site is underlined; inconsistent G and C stretches are shaded.

```
cgacagcacacacagccgaaaagcgcttcttctcatcgccttcgccttcaaaactttgaagcaccaatccca    72
accagaaaaa ATG TGC GAC GAA GAA GTT GCT GCT CTC GTG GTG GAC AAT GGG TCC     127
            M   C   D   E   E   V   A   A   L   V   V   D   N   G   S      15
GGA ATG TGC AAG  gcgagttttgggatttggacttattattatgtacaattgggaaccatcaaaatgg   194
 G   M   C   K                                                            19
gcaaaagcaatgcttttgatatgctaaatctcaatttgggcattggccacatttttacagtgaaattcaaac   266
aaatttttgatcaaatcaaccgaaaataatttttccgaatatttttaaaataaatcaaatttcataagaaaat   338
gcccactcgggcaaaatgtctcctcattattggaaaaagaaaaaagcaaatccccccccccggggtggaaat   410
tctccgaattctttgagcacaacgaatgtttttgcttcgccaacaaaaaaagcgggcgatgggggggggga   482
gagaatgttttatttgtgactcacacttttgaccgacttttctcttcagaatgttacaattccaagaagagg   554
gagaaaaaggggggaaaagaataaaaagaaaacaaaaattggctaattaaaattggattgaatttccgcttta   626
attcagGCC GGT TTC GCG GGC GAT GAC GCG CCC CGC GCG GTG TTT CCG TCC ATT       680
       A   G   F   A   G   D   D   A   P   R   A   V   F   P   S   I       35
GTG GGC CGC CCC CGT CAT CAG GGC GTC ATG GTC GGC ATG GGA CAG AAG GAC TCG    734
 V   G   R   P   R   H   Q   G   V   M   V   G   M   G   Q   K   D   S     53
TAC GTG GGA GAC GAG GCG CAG TCC AAG CGT GGT ATT CTG ACG CTG AAG TAC CCG    788
 Y   V   G   D   E   A   Q   S   K   R   G   I   L   T   L   K   Y   P     71
ATT GAG CAC GGC ATC GTC ACC AAC TGG GAC GAC ATG GAG AAG ATC TGG CAC CAC    842
 I   E   H   G   I   V   T   N   W   D   D   M   E   K   I   W   H   H     89
ACC TTC TAC AAC GAG CT gtgagtggtggtgccatcgtcagattggcattggcactcgccgtcggtc   908
 T   F   Y   N   E                                                        94
acttggccaatccgtccctttaagG CGT GTC GCC CCC GAG GAG CAC CCG GTG CTG CTC      966
                         L   R   V   A   P   E   E   H   P   V   L   L     106
ACA GAG GCC CCG CTG AAC CCG AAG GCC AAC AGA GAA AAG  gtctatctgccaaatggaa  1024
 T   E   A   P   L   N   P   K   A   N   R   E   K                       119
cagtgaaactgacggaatctccccttttcagATG ACC CAA ATC ATG TTC GAG ACG TTC AAC   1085
                               M   T   Q   I   M   F   E   T   F   N     129
ACG CCG GCC ATG TAC GTG GCC ATC CAG GCC GTG CTG TCC CTG TAC GCC TCC GGC   1139
 T   P   A   M   Y   V   A   I   Q   A   V   L   S   L   Y   A   S   G     147
CGT ACC ACC GGT CTC GTG TTG GAC TCC GGC GAT G gtatgcagccaaatcgttggcataat  1199
 R   T   T   G   L   V   L   D   S   G   D                               158
atcatagctatcaatctccatgcacagGT GTC ACC CAC ACC GTG CCC ATC TAC GAG GGT     1258
                             G   V   T   H   T   V   P   I   Y   E   G     169
TAC GCC CTG CCC CAC GCC ATC CTC CGT TTG GAC TTG GCC GGC CGT GAC CTC ACT   1312
 Y   A   L   P   H   A   I   L   R   L   D   L   A   G   R   D   L   T     187
GAC TAC CTG ATG AAG ATC CTC ACT GAG CGT GGT TAC TCC TTC GTG ACC ACG GCC   1366
 D   Y   L   M   K   I   L   T   E   R   G   Y   S   F   V   T   T   A     205
GAG CGT GAG ATC GTC CGT GAC ATC AAG GAG AAG TTG TGC TAC GTG GCA CTC GAC   1420
 E   R   E   I   V   R   D   I   K   E   K   L   C   Y   V   A   L   D     223
TTC GAG CAG  gtcggacatttggccaatttgatgccaaaattacccttcctcccgtcttggttcagGAA  1488
 F   E   Q                                                           E     227
ATG GCC ACT GCC GCC TCC TCC TCC TCG CTG GAG AAG AGC TAC GAA CTG CCC GAC   1542
 M   A   T   A   A   S   S   S   S   L   E   K   S   Y   E   L   P   D     245
GGC CAA GTG ATC ACC GTC GGC AAC GAG CGC TTC CGT TGC CCA GAG GCG CTG TTC   1596
 G   Q   V   I   T   V   G   N   E   R   F   R   C   P   E   A   L   F     263
CAG  gtgagtgcccggcggagggcgcgtgcctatggcaacggcacaatttgatttggctctgttcagCCG  1665
 Q                                                               P     265
TCC TTC ATC GGC ATG GAG TCG GCC GGC ATC CAC GAG TCC ACC TAC AAC TCG ATC   1719
 S   F   I   G   M   E   S   A   G   I   H   E   S   T   Y   N   S   I     283
ATG AAG TGC GAC ATC GAC ATC CGT AAG GAC CTG TAC GCC AAC AAT GTG CTG TCC   1773
 M   K   C   D   I   D   I   R   K   D   L   Y   A   N   N   V   L   S     301
GGC GGC ACC ACC ATG TAC CCG GGC ATT GCT GAC CGC ATG CAG AAG GAG ATG ACC   1827
 G   G   T   T   M   Y   P   G   I   A   D   R   M   Q   K   E   M   T     319
GCG CTG GCC CCG TCC ACC ATG AAG ATC AAA ATC ATC GCT CCC CCG GAG CGC AAG   1881
 A   L   A   P   S   T   M   K   I   K   I   I   A   P   P   E   R   K     337
TAC TCC GTC TGG ATC GGC GGC TCC ATC CTC GCC TCC CTG TCC ACC TTC CAA CAG   1935
 Y   S   V   W   I   G   G   S   I   L   A   S   L   S   T   F   Q   Q     355
AT gtgagttgtggttgcattgggctgatctgacccaccaatgcttttccccctctgatcagttgcccaatt  2006
agctgatcattgtccgttgttttccccctctgagcactcttttgccctttctcagG TGG ATC AGC     2072
                                                       M   W   I   S       359
AAG CAG GAG TAC GAC GAG TCG GGC CCG TCC ATT GTC CAC CGC AAG TGC TTC TGA   2126
 K   Q   E   Y   D   E   S   G   P   S   I   V   H   R   K   C   F   *     377
gcagcagctgcacgaccgcacggaggaggaggagaagaccaacgaagccgggacactcgccatccgccgacc  2198
cattcgccacccccccaccaccatctgtttttttgctgaaaattatgcaaaaatatttttgaataaaatctatg  2270
aacaaaaaaaaaaaaaa                                                       2286
```

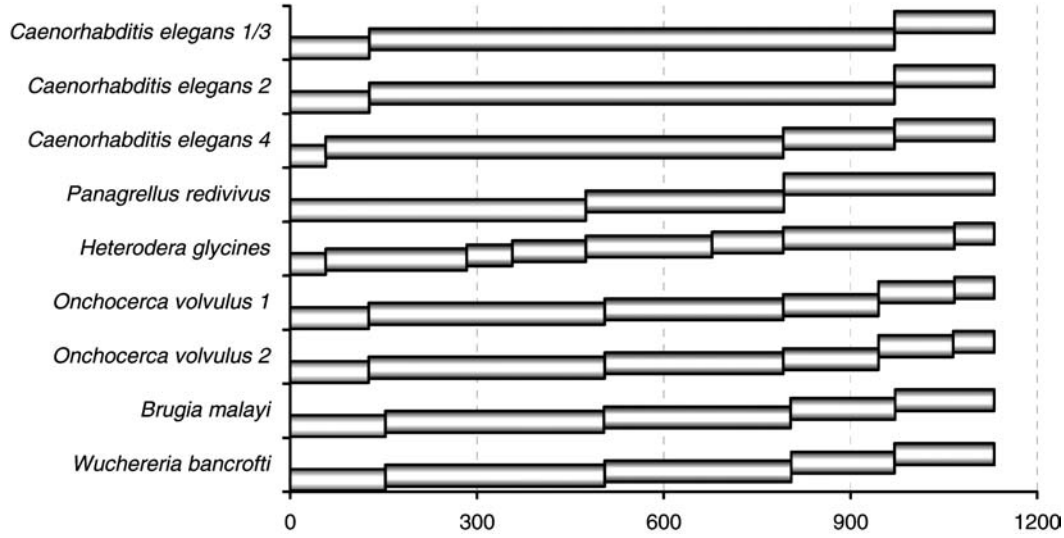

Figure 4. Diagram of exon composition of nematode actin genes. Axis numbers correspond to coding sequence nucleotides. Actin genes are schematically shown as combinations of exons of proper length. Intron locations are illustrated by shifts between exons.

any plant-parasitic nematode. The similar organization of proteins from the superfamily makes the acquired information about actin genes useful for discovery of other members. Because of the extreme biases of these genes in cyst nematodes, the established preferred codon usage is of direct practical relevance in minimizing degeneracy of PCR primers and optimizing conditions. Because actin is a common housekeeping protein, it is often used as an internal standard gene for comparison with target genes in Real-Time PCR. Thus, the actin gene sequence from *H. glycines* was successfully used in our laboratory for estimation of expression levels of several genes during development and under stress conditions.

Actin gene in cyst nematodes

Comparison of 3 genes encoding actin from *H. glycines* and *G. rostochiensis* and partial sequences (exons 2 through 8) from 6 other *Heterodera* spp. unequivocally indicates the high similarity of actin genes from cyst nematodes. Actin amino acid identity for the cyst nematodes was 99–100%, and all sequences contained 4 diagnostic amino acid residues (L-152, V-202, I-268, and M-318). Actin nucleotide identity was also high (94–99%), and the genes were thus similar in codon usage bias. All cyst nematode actin genes showed a predominance of G or C at the third position, with a very low usage of A; the GC3 value for tested cyst nematodes varied from 0.83 to 0.87. Six examined introns had the identical locations, similar lengths (seventh the longest, third–fourth–fifth the shortest) and moderate AT proportion within individual introns (the average $<50\%$).

Phylogenetic analysis based on coding nucleotide sequences of actin genes yielded results similar to those inferred from rDNA sequences, e.g., for cyst nematode internal transcribed spacer (ITS) sequences (Subbotin et al., 2000). The positions of *Panagrellus* and *Plectus* on the tree were not resolved. The unexpected clustering of some taxa in the actin nucleotide maximum parsimonious tree possibly resulted from high mutational saturation or nucleotide frequency biases (or both) for different nematode groups. The variability of nucleotide sequences within introns along with a particular advantage of unambiguous alignment of coding regions might be used for population studies on cyst nematodes.

Actin gene in plant-parasitic nematodes

To study more distantly related plant-parasitic nematodes, we examined partial sequences for actins from 4 root-knot nematodes: *Meloidogyne javanica* (AF532605), *Meloidogyne arenaria, Meloidogyne incognita*, and *Meloidogyne hapla* (expressed sequence tag [EST] database). Fragments with unequivocal ORF from the EST database were translated to protein sequences, and in areas of

interest, we found that none of the 4 diagnostic residues were identical to those in cyst nematode actins. Instead, these residues were V-152 (for *M. incognita*, *M. arenaria*), I-152 (for *M. hapla*, *M. javanica*), and T-202, L-268, and I-318 (all *Meloidogyne* spp.), typical for free-living and animal-parasitic nematodes. Partial messenger RNA sequence from *M. javanica* (630 bp) had only 77% identity to the corresponding sequence from *H. glycines*; and corresponding GC3 values for the fragments were 0.41 and 0.87. Partial genomic sequence from *M. javanica* (AF532604) included 5 introns, 2 of which were placed identically to the fourth (159/1) and sixth (264/3) introns in actin from *Heterodera* spp., the latter the most conserved among nematodes; positions of 3 other introns in *M. javanica* actin were matchless.

The actin gene structures of these 2 groups of morphologically similar and taxonomically related nematodes—the cyst nematodes (*Heterodera* and *Globodera* spp.) and the root-knot nematodes (*Meloidogyne* spp.)—were conspicuously different. Nematodes of the family Hoplolaimidae, according to classical morphological classification, are related to both cyst and root-knot nematodes (Siddiqi, 2000). A preliminary study on actin genes from the Hoplolaimidae revealed that 2 genes possessed diagnostic amino acid residues typical for cyst nematodes, whereas 1 gene had residues typical for root-knot and other nematodes. All 3 genes had codon usage biases typical of cyst nematodes, with a high GC3 value (about 0.8), and intron locations were identical to those from cyst nematodes (unpublished data).

Actin gene in nematodes

Despite the very high amino acid similarity among nematode actins, nucleotide coding sequences of actins from cyst nematodes were strikingly different from those of free-living and animal-parasitic nematodes. Indeed, the best matches in nucleotide sequences for actin CDSs from *H. glycines* and *G. rostochiensis* were found among chordates (bony-fishes, lancelets, rabbit, lampreys, birds, etc.), whereas nucleotide sequences of filarial nematodes (with 98% amino acid identity) trailed behind a long list of species including plants and fungi. This could be explained by the difference in codon usage between cyst nematodes and other nematodes, which is characteristic of genome. Actins, as well as most protein-coding sequences of *C. elegans* and filarial nematodes, are AT rich with higher frequencies of codons ending in an A or a T base (Hammond and Bianco, 1992; Unnasch et al., 1992; Hammond, 1994; Ellis et al., 1995; Fadiel et al., 2001). Introns in filariids and *C. elegans* were shown to be even more AT rich (at least 70%) than the coding region (Blumenthal and Steward, 1997; Saverimuttu et al., 2000; Unnasch and Williams, 2000).

Comparison of the exon–intron organization of nematode actin genes revealed that the first (19/3) and sixth (264/3) introns in *H. glycines* actin are situated in the same positions as reported for the first and second introns of *C. elegans* Act4. The latter appears to be most typical for nematode actin genes, coinciding with intron positions in *P. redivivus*, *O. volvulus*, and *S. digitata* actin genes and being close to intron position in *B. malayi* and *W. bancrofti*. The survey of actin gene introns beyond the phylum Nematoda demonstrated that the positions of the first (19/3) and the last (356/2) introns coincide with the positions of the first and last actin gene introns in many higher plants, including *Glycine max* (the major host of *H. glycines*). The fifth (226/3) intron of *H. glycines* has not been reported and occupies a novel site in the catalogue of more than 50 intron positions known for actin genes from published (Bagavathi and Malathi, 1996) or Internet-based (ExInt) sources.

In conclusion, we have cloned and characterized the actin gene from 2 cyst nematodes, major pests of soybean and potato. Comparison of sequences with those from other nematodes revealed that the described features are characteristic for a wide range of cyst nematodes but differ from other nematodes, including plant-parasitic nematodes in the family Meloidogynidae, traditionally placed close to the Heteroderidae because of morphological similarity. The unambiguous alignment of coding nucleotide sequences and identical location of introns with variable intron sequences make actin a convenient marker for phylogenetic study of cyst nematodes. The simultaneous usage of different phylogenetic approaches, such as amino acid sequence, coding nucleotide sequence, codon usage, intron location, may help to resolve some of the taxonomic problems of the large group of living nematodes (Blaxter et al., 1998).

ACKNOWLEDGMENTS

Support for S.A.S. from the CLO-Agricultural Research Centre, Merelbeke, Belgium, is gratefully acknowledged. We thank Dr. Eric Davis (North Carolina State University) for the cDNA library from *H. glycines* and Dr. Geert Smant (Laboratory of Nematology, Wageningen University and Research

Centre) for the cDNA from *G. rostochiensis*. We gratefully acknowledge M. Romero, U. Scholz, D. Sturhan, and W. Wouts for supplying nematode samples. We thank Dr. Lynn Carta, USDA, for assistance in systematic and phylogenetic analysis.

Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

LITERATURE CITED

Bagavathi, S., and R. Malathi. 1996. Introns and protein revolution–an analysis of the exon/intron organisation of actin genes. FEBS Letters 392:63–65.

Baldauf, S. L., A. J. Roger, I. Wenk-Siefert, and W. F. Doolittle. 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. Science 290: 972–977.

Blaxter, M. L., P. De Ley, J. R. Garey, L. X. Liu, P. Scheldeman, A. Vierstraete, J. R. Vanfleteren, L. Y. Mackey, M. Dorris, L. M. Frisse, J. T. Vida, and W. K. Thomas. 1998. A molecular evolutionary framework for the phylum Nematoda. Nature 392:71–75.

Blumenthal, T., and K. Steward. 1997. Chapter 6. RNA processing and gene structure. Pages 117–145 *in* D. L. Riddle, T. Blumenthal, B. J. Meyer, and J. R. Priess, eds. *C. elegans* II. Cold Spring Harbor Laboratory Press, Plainview, New York.

Bork, P., C. Sander, and A. Valencia. 1992. An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins. Proceedings of the National Academy of Sciences of the United States of America 89:7290–7294.

Copley, R. R., J. Schultz, C. P. Ponting, and P. Bork. 1999. Protein families in multicellular organisms. Current Opinion in Structural Biology 9:408–415.

De Loof, A., J. Vanden Broeck, and I. Janssen. 1996. Hormones and the cytoskeleton of animals and plants. International Review of Cytology 166:1–58.

Donnelly, S. M., D. J. Sullivan, D. B. Shanley, and D. C. Coleman. 1999. Phylogenetic analysis and rapid identification of *Candida dubliniensis* based on analysis of ACT1 intron and exon sequences. Microbiology 145:1871–1882.

Ellis, J., D. A. Morrison, and B. Kalinna. 1995. Comparison of the patterns of codon usage and bias between *Brugia*, *Echinococcus*, *Onchocerca* and *Schistosoma* species. Parasitology Research 81: 388–393.

Fadiel, A., S. Lithwick, M. Q. Wanas, and A. J. Cuticchia. 2001. Influence of intercodon and base frequencies on codon usage in filarial parasites. Genomics 74:197–210.

Files, J. G., S. Carr, and D. Hirsh. 1983. Actin gene family of *C. elegans*. Journal of Molecular Biology 164:355–375.

Hammond, M. 1994. Codon usage and gene organization in *Brugia*. Parasitology Research 80:173–175.

Hammond, M. P., and A. E. Bianco. 1992. Genes and genomes of parasitic nematodes. Parasitology Today 8: 299–305.

Holland, J. B., S. J. Helland, N. Sharopova, and D. C. Rhyne. 2001. Polymorphism of PCR-based markers targeting exons, introns, promoter regions, and SSRs in maize and introns and repeat sequences in oat. Genome 44:1065–1076.

Hurley, J. H. 1996. The sugar kinase/heat shock protein 70/actin superfamily: implications of conserved structure for mechanism. Annual Review of Biophysics and Biomolecular Structure 25:137–162.

Kovaleva, E. S., A. G. Yakovlev, E. P. Masler, and D. J. Chitwood. 2002. Human proprotein convertase 2 homologue from a plant nematode: cloning, characterization, and comparison with other species. FASEB Journal 16:1099–1101.

Krause, M., M. Wild, B. Rosenzweig, and D. Hirsh. 1989. Wild-type and mutant actin genes in *Caenorhabditis elegans*. Journal of Molecular Biology 208: 381–392.

Lee, J. S., and M. C. Gye. 2001. Use of beta-actin gene intron 2 as a phylogenetic marker in fish taxonomy. DNA Sequence 12:71–76.

Masler, E. P., E. S. Kovaleva, and S. Sardanelli. 1999. Comparison of FaRP immunoreactivity in free-living nematodes and in the plant-parasitic nematode *Heterodera glycines*. Annals of the New York Academy of Sciences 897:253–263.

Ohresser, M., P. Borsa, and C. Delsert. 1997. Intron-length polymorphism at the actin gene locus mac-1: a genetic marker for population studies in the marine mussels *Mytilus galloprovincialis* Lmk. and *M. edulis* L. Molecular Marine Biology and Biotechnology 6: 123–130.

Saverimuttu, J. K. C., E. H. Karunanayake, N. V. Chandrasekharan, and S. M. T. Jayasena. 2000. Molecular characterization of the actin gene of the filarial parasite *Wuchereria bancrofti*. International Journal for Parasitology 30:119–124.

Siddiqi, M. R. 2000. Tylenchida: Parasites of Plants and Insects, 2nd ed. CABI Publishing, New York. 833 pp.

Subbotin, S. A., A. Vierstraete, P. De Ley, J. Rowe, L. Waeyenberge, M. Moens, and J. R. Vanfleteren. 2001. Phylogenetic relationships within the cyst-forming nematodes (Nematoda, Heteroderidae) based on analysis of sequences from the ITS regions of ribosomal DNA. Molecular Phylogenetics and Evolution 21:1–16.

Sutherland, J. D., and W. Witke. 1999. Molecular genetic approaches to understanding the actin cytoskeleton. Current Opinion in Cell Biology 11: 142–151.

Swofford, D. L. 1998. PAUP. Phylogenetic analysis using parsimony and other methods. Version 4. Sinauer Associates, Sunderland, Massachusetts. 128 pp.

Tardieux, I., I. Baines, M. Mossakowska, and G. E. Ward. 1998. Actin-binding proteins of invasive malaria parasites and the regulation of actin polymerization by a complex of 32/34-kDa proteins associated with heat shock protein 70kDa. Molecular and Biochemical Parasitology 93:295–308.

Unnasch, T. R., C. R. Katholi, and L. M. Coate. 1992. *Onchocerca volvulus*: frequency of codon usage. Experimental Parasitology 75:457–459.

Unnasch, T. R., and S. A. Williams. 2000. The genomes of *Onchocerca volvulus*. International Journal for Parasitology 30:543–552.

Voigt, K., and J. Wostemeyer. 2000. Reliable amplification of actin genes facilitates deep-level phylogeny. Microbiological Research 155:179–195.

Zeng, W, and J. E. Donelson. 1992. The actin genes of *Onchocerca volvulus*. Molecular and Biochemical Parasitology 55:207–216.