Journal of Nematology 37(4):422–428. 2005.

Characterization of Several *Heterodera glycines* mRNA that Encode Small Proteins with Putative Signal Peptides

Mark L. Tucker, Ping Xue, Anjana Raina, Mindy L. Ehrenfried, Mehar Asif, Vanessa K. Thai

Abstract: Two subtraction libraries were prepared from RNA extracted at early and late stages in the development of soybean cyst nematodes (SCN), *Heterodera glycines,* in soybean roots. The cDNA from inoculated roots were subtracted with cDNA prepared from non-inoculated roots and SCN eggs, and 384 clones from each library were sequenced. BLAST searches revealed that 191 of the cDNA in the late library were most probably of nematode origin. Alignment of the 191 sequences produced 28 unigenes and 1 singlet. The size of the transcripts for the nematode genes was confirmed by RNA blot hybridization. Thirteen SCN transcripts were selected for further study because they included short open reading frames encoding predicted proteins of <20 kDa with signal peptides at their amino-terminus. Ten of the 13 encode predicted peptides <10 kDa. Although most of the 13 transcripts were fairly abundant in the SCN dbEST, most were of unknown function based on BLAST similarities. Nevertheless, several had characteristics common to anti-microbial peptides, and in situ hybridization indicated that three of the selected transcripts were expressed in the female reproductive system.

Key words: anti-microbial, *Heterodera glycines,* reproductive system, peptides, soybean cyst nematode, subtraction library.

Small peptides are difficult to visualize on standard acrylamide gels because they are unresolved at the leading edge of the running buffer. Moreover, standard procedures for cDNA library construction include a size fractionation of the cDNA to reduce the number of clones for partial length RNA and to eliminate adaptor sequences. These procedural steps unintentionally reduce the likelihood of cloning short but full-length transcripts. This and a natural tendency to assume that short peptides and cDNA fragments were artifacts of the isolation procedure obscured the importance of small peptides in cell biology. Nevertheless, evidence indicates that small peptides synthesized in the nematode may serve several functions including neuropeptides (Brownlee et al., 2000; Li et al., 1999b), anti-microbial agents (Boman, 2003; Bulet et al., 2004), and other internal and externally secreted developmental signals (Bird, 2004; Davis et al., 2004).

Neuropeptides are small signaling proteins expressed in the nervous system of most higher organisms (Li et al., 1999a). They are typically expressed as a pre-protein that is processed to activate the signaling peptide. They have been particularly well studied in *C. elegans* where 92 genes have been identified that include conserved sequence motifs described for neuropeptides characterized in many organisms (Nathoo et al., 2001). Anti-microbial peptides are common to micro-organisms, plants, and animals (Bulet et al., 2004). Anti-microbial peptides are typically 15 to 45 amino acids long and possess a net positive charge (Boman, 2003). Boman (2003) suggested that most species encode between 15 and 40 genes for anti-microbial peptides.

It has been proposed that plant-parasitic nematodes secrete small peptides into the host where they alter host cell biology (Davis et al., 2004; Gao et al., 2003). Parasitism of plants by cyst nematodes is a complex process requiring successful penetration and migration of the nematode into the root to the vascular bundle, where the nematode induces the formation of a feeding structure, i.e., syncytium. The signals that evoke changes in the host cell to form the feeding structure could take the form of chemicals (e.g., cytokinins) or proteins (Bird, 2004). Plant-parasitic nematodes may have exploited small peptides as signals and antagonists as a means to co-opt the host into forming and maintaining the nematode feeding structure (Bird, 2004; Davis et al., 2004; Williamson and Gleason, 2003).

As a means to identify transcripts of genes up-regulated by infection of soybean roots by the soybean cyst nematode (SCN) *Heterodera glycines,* we prepared two subtraction libraries (Duguid and Dinauer, 1990) from SCN-infected soybean roots. One library was made with RNA from 1, 2, and 4 days post-inoculated (dpi) roots and a second from 8, 12, and 20 dpi roots. We sequenced 384 clones from each library and arrayed them together onto a membrane macroarray. We found that 50% of the late library consisted of SCN transcripts, and most of the SCN clones were cDNA fragments for transcripts that included ORFs for small proteins with putative amino-terminus signal peptides. Northern blots and 5′ and 3′ RACE (rapid amplification of cDNA ends) confirmed that many of these cDNA fragments represented small mRNA expressed during SCN development in the soybean root and not in eggs or second-stage juveniles (J2).

Materials and Methods

Nematode and plant material: The SCN population NL1-RH, which interacts with soybean differentials in the manner of race 3, has been maintained on greenhouse-grown soybeans at the U.S. Department of Agriculture, ARS, Beltsville, Maryland, and can be obtained

Received for publication 8 April 2005.

Soybean Genomics and Improvement Laboratory, U.S. Department of Agriculture, Building 006, BARC-west, 10300 Baltimore Ave. Beltsville, MD 20705.

The authors thank M. MacDonald and B. Matthews for discussions about a variety of topics relating to nematode research. GenBank accession numbers: Hg-Con1, ay853176; Hg-Con2, ay853178; Hg-Con3, ay853175; Hg-Con4, ay853173; Hg-Con5, ay853174; Hg-Con6, ay853177.

E-mail: tuckerm@ba.ars.usda.gov

This paper was edited by David Bird.

from the SCN Stock Center (T. Niblack, University of Illinois, Champaign-Urbana, IL). The nematode inoculum was prepared as previously described (Matthews et al., 2003). *Glycine max* cv Williams 82 seeds were germinated in Perlite (Geiger, Harleysville, PA) in the greenhouse and after 2 weeks seedlings were washed free of Perlite, combined into groups of five seedlings, and inoculated by pipeting 2,500 J2/seedling. The roots were then sprinkled with autoclaved moist Perlite and covered with a moist paper towel. The seedlings were kept in a moist environment in a growth chamber with 15 hours of light per 24-hour cycle. After the desired incubation interval, one of five seedlings in each group was collected for acid fuchsin staining to monitor nematode development and number of nematode infections (Byrd et al., 1983). The roots from the remaining four plants were frozen in liquid nitrogen and stored at −70°C.

Subtraction library: RNA was extracted from roots as previously described (Koehler et al., 1996) from non-inoculated plantlets at the beginning of the treatment period (0 days post-inoculated [dpi]) and then at 1, 2, 4, 8, 12, and 20 dpi from non-inoculated and SCN inoculated plantlets. RNA from eggs and J2 were extracted using an RNeasy plant mini kit from Qiagen (Carlsbad, CA). Total RNA was combined in equal portions from 1, 2, and 4 dpi non-inoculated roots (E−); 1, 2, and 4 dpi SCN inoculated roots (E+); 8, 12, and 20 dpi non-inoculated roots (L−); and 8, 12, and 20 dpi SCN inoculated roots (L+). Five µg of E−, E+, L−, L+, and SCN egg RNA were treated separately with RNase-free DNase (Promega, Madison WI) and first-strand cDNA synthesis completed with 1 µg of each RNA mixture using the SMART PCR cDNA Synthesis kit (BD Biosciences, Palo Alto, CA). Two subtraction libraries were then prepared using the Clontech PCR-Select cDNA Subtraction Kit (BD Biosciences). The tester PCR products for E+ and L+ were subtracted with the E− and L− PCR products, respectively, plus SCN egg PCR products. The subtracted PCR products were then cloned into pGEM-T (Promega), and 1,920 individual colonies from each library were collected into 96-well microtiter plates.

Cloned cDNA inserts were amplified with M13 forward and reverse primers and PCR products prepared for sequencing by treatment with shrimp alkaline phosphatase and exonuclease I for 60 minutes at 37°C. Cleaned PCR products were sequenced using the ABI Big DyeTM Terminator Cycle Sequencing Kit and run on the ABI Prism 3100 Genetic Analyzer (Perkin-Elmer, Applied Biosystems, Foster City, CA). Inserts were sequenced in both directions. To obtain full-length sequence for selected clones, 5′ and 3′ RACE was performed using the GeneRacer kit (Invitrogen, Carlsbad, CA).

RNA blots: RNA was extracted from roots, eggs, and J2 as described above and approximately 5 µg loaded into each lane of six gels that were then transferred to Hybond N nylon membranes (Amersham, Piscataway, NJ) by capillary transfer (Sambrook et al., 1989). Nematode cDNA clones were labeled by incorporation of ^{32}P dCTP into a PCR reaction. Hybridization conditions for RNA blots were 60% formamide, 5× SSPE, 2× Denhardt's, 0.1% SDS, and 150 µg/ml denatured salmon sperm DNA (Sambrook et al., 1989) at 50°C overnight. The final wash of the blots was 0.1× SSPE, 0.1% SDS, and 0.005% sodium pyrophosphate at 55°C for 60 minutes. Blots were exposed to Biomax X-ray film with intensifying screens at −70°C. After exposure to film the blots were either stored for several weeks to allow the radioactivity to decay or immediately washed free of probe by submerging the blot in 0.1× SSPE plus 0.1% SDS at 100°C and then allowing the submerged blot to cool slowly with gentle shaking.

In situ hybridization: Williams 82 soybean seeds were surface-sterilized with 95% ethanol for 3 minutes followed by 10% bleach for 10 minutes and germinated on 1.5% agar in the dark for 3 days. Two root tips per plate were cut and grown on 1.5% Noble agar containing Gamborg's B5 medium plus organics and 2% sucrose. After 2 days at 25°C five SCN females from an earlier cultured plate were crushed to release eggs alongside each root. Females were collected 15 to 20 days post inoculation and fixed overnight in 2% freshly prepared paraformaldehyde (PFM). Nematodes were centrifuged to the bottom of the tube and PFM diluted to 0.2% with phosphate buffered saline, and the nematode sample was stored at −70°C.

In situ hybridization was performed essentially as described by de Boer et al. (1998). The frozen nematodes were thawed in 0.2% PFM and 0.1× M9 buffer and each female cut in half with a razor blade. The RNA probe was prepared using a UTP-digoxigenin labeling kit (Roche Diagnostics, Indianapolis, IN) with T3 or T7 polymerase for sense and antisense probes, respectively. The probes were reduced to an average length of 125 bp by alkaline hydrolysis (Sambrook et al., 1989).

Computer analysis: Sequences were trimmed of vector and adaptors manually in the editor window of the GCG Wisconsin Package (Accelrys, San Diego, CA) software and all the sequences checked for remaining vector with a FASTA search of the file with the pGEM-T vector sequence. A BLASTN search of the *H. glycines* GenBank dbEST was used to identify sequences similar to the subtraction library cDNA, and all similar sequences were entered into the GCG assembly program for alignment and generation of contiguous consensus sequences. Open reading frames, signal peptides, and isoelectric points were determined with the GCG applications Map, SPScan, and Isoelectric, respectively, using the default settings.

Results

Sequence analysis: Two subtraction libraries were prepared with RNA from 1, 2, and 4 days or 8, 12, and 20 dpi roots and subtracted with SCN egg RNA and RNA from non-inoculated roots incubated for the same number of days. From each library, 384 clones were sequenced from both ends. BLASTN and TBLASTX searches of the public nr, dbEST, and PIR databases, respectively, were completed for each clone using the GCG default settings with filtering on to reduce alignment with low-complexity sequences. The clones were grouped by origin (plant, nematode, fungal/bacterial, or unknown) based on an E-score of less than $1 \times e^{-20}$ with an emphasis on hits in non-EST databases. A cDNA was assigned a soybean origin if the sequence best matched multiple plant sequences from any plant species and an SCN origin if it best matched a sequence in a database for any nematode species. A similar criterion was used to identify a clone as having a fungal or bacterial origin. Sequences that did not match any sequence in the searched databases were labeled as unknown. Based on these criteria, none of the 384 clones from the 1-to-4 dpi library were of SCN origin. However, in the 8-to-20 dpi library at least 191 cDNA had high-sequence identities with a sequence in the SCN database or a database for another nematode species. The 191 cDNA were merged with all of the similar *H. glycines* GenBank nr and dbEST sequences to produce 28 contiguous consensus sequences (unigenes) and 1 singlet that had no highly similar EST in the *H. glycines* database (Table 1). One contig, Hg-Con2, included 99 subtraction library clones and another five contigs (Hg-Con1, 3, 4, 5, and 6) included between 9 and 16 cDNA from the 384 clones sequenced (Table 1). Interestingly, all six of these contigs included ORF for peptides with putative amino-terminus signal peptides that indicate synthesis on the endoplasmic reticulum and possible secretion outside the cell (von Heijne, 1983). Moreover, five of the six contigs encoded small proteins of 13.2, 6.5, 5.6, 4.2, and 3.3 kDa with no clear function based on primary-sequence comparisons (Table 1). Further analysis of the SCN clones in the late subtraction library revealed another eight unigenes that included ORFs encoding small proteins with amino-terminus signal sequences (Table 1). The isoelectric points for the 13 mature peptides varied from 4.0 to 10.5 (Table 1).

Determination of transcript length and expression patterns: RNA blots were completed to confirm the predicted size of the 13 selected transcripts and define the temporal expression patterns for the transcripts (Fig. 1). For comparison, in addition to the 13 transcripts selected from the subtraction library, Figure 1 includes expression profiles for several better-characterized SCN and soybean transcripts (Gao et al., 2003; Matthews et al., 2003). The soybean elongation factors 1a and 1b (Gm-EF1a and Gm-EF1b, Fig. 1) were used because both are constitutively expressed in soybean roots and span the size range of the mRNA being probed. The SCN-infected roots used for the RNA blots shown in Figure 1 had approximately 400 infections/root at 2 dpi. At 20 dpi there were approximately 90 swollen females/root. Although only 20% of the nematodes that infect the root matured into adult females, the hybridization signal for the constitutively expressed elongation factor is considerably greater at later stages of development because the nematode has grown over the 20 dpi. Comparison of the RNA blots demonstrates that the 13 selected transcripts are not expressed at detectable levels in eggs or J2 but are developmentally expressed as the nematode grows in the soybean root (Fig. 1). Moreover, based on their electrophoretic migration in the agarose gel, the length of the 13 transcripts is close to that predicted from the derived consensus sequences.

Because the consensus sequences were derived from EST data that sometimes includes artifacts, primers were prepared and 5′ and 3′ RACE completed on the first six unigenes in Table 1 to confirm their full-length sequence and ORF. Few differences were found between the ORF of the derived consensus sequences and the confirmed sequences for these cDNA.

In situ hybridization: In situ hybridization was performed successfully on four of the 13-selected small mRNA (Fig. 2). Hg-Con2 hybridization first appears in the anterior half of an immature female (Fig. 2A) and then in a tubular structure in a later stage of female development that in this micrograph has broken away from its normal attachments but still encloses a few maturing eggs (Fig. 2B). Hg-Con3 had a similar pattern of expression in a tubular structure in the anterior half of a maturing female (Fig. 2C) and again in a tubular structure with enclosed eggs in the posterior half of an adult female (Fig. 2D). Hg-Con6 has a similar expression pattern as the other two except that we did not observe expression in the anterior half of immature females (Fig. 2E,F). Hg-Con5 hybridization was different from the other three (Fig. 2G). Again the hybridization was associated with a mostly tubular structure that didn't appear to contain eggs. Also included in Figure 2 is an in situ hybridization for Hg-SYV46 (Gao et al., 2003), which is expressed in the dorsal esophageal gland of an adult female (Fig. 2H). The latter in situ hybridization was included as a control because it is a better-characterized gene whose organ-specific expression has been published previously by others (Gao et al., 2003).

Discussion

Microscopic examination of roots stained for SCN indicated that by 4 dpi a syncytium is evident and at 8

TABLE 1. *H. glycines* unigenes identified in a late subtraction library and properties derived from the consensus sequences for each unigene. The first six sequences were confirmed by obtaining full-length clones.

Name[1]	Accession number[2]	Clones in Sub-lib[3]	EST in Hg database[4]	Length[5] (bp)	Poly(A) tail[6]	Full ORF[7]	Signal pep[8]	Mature pep (kDa)[9]	Isoelectric point[10]	Most similar protein in uniprot database (U)[11] or hidden Markov model (M)[12]
Hg-Con1*	ay853176	12	21	367	yes	yes	yes	3.3	7.7	none
Hg-Con2*	ay853178	99	61	456	yes	yes	yes	6.5	6.1	none
Hg-Con3*	ay853175	10	62	494	yes	yes	yes	5.6	11.5	none
Hg-Con4	ay853173	10	6	1172	yes	yes	yes	32.6	8.2	esophageal gland, *H. glycines* (U)
Hg-Con5*	ay853174	9	43	359	yes	yes	yes	4.2	10.4	none
Hg-Con6*	ay853177	14	6	699	yes	yes	yes	13.2	10.3	eggshell protein, *S. mansoni* (U)
Hg-Con7	ca940212	6	61	1937	NA[13]	NA	NA	NA	NA	28S rRNA[14]
Hg-Con8	af498244	4	10	2048	yes	yes	yes	71.3	5.5	C-type lectin (U&M)
Hg-Con9	co036789	3	0	174	no	no	NA	NA	NA	hypothetical protein, *C. elegans* (U)
Hg-Con10	ck394243	2	73	1175	yes	yes	yes	25.1	7.2	collagen (M)
Hg-Con11*	co036615	2	55	320	yes	yes	yes	3.4	10.5	none
Hg-Con12	cb279343	2	11	1116	NA	NA	NA	NA	NA	18S rRNA
Hg-Con13	co036623	2	1	856	no	no	Na	NA	NA	dynein (U), WD domain (M)
Hg-Con14*	cb374836	2	20	445	yes	yes	yes	5.8	6.8	none
Hg-Con15*	ca940933	1	4	515	no	yes	yes	9.2	10	none
Hg-Con16	cb279403	1	2	639	no	no	NA	NA	NA	hypothetical protein, *C. elegans* (U)
Hg-Con17	ca940661	1	3	844	no	no	NA	NA	NA	ubiquitin C-term hydrolase (U&M)
Hg-Con18	cb378367	1	8	710	yes	yes	no	22.4	8.9	ATP synthase (U&M)
Hg-Con19	ca939432	1	13	1460	yes	no	NA	NA	NA	proline-rich proteins
Hg-Con20	ck350214	1	12	1491	yes	yes	no	47.5	6.8	enolase (U&M)
Hg-Con21	cb826100	1	1	875	no	no	NA	NA	NA	hypothetial protein *C. elegans* (U)
Hg-Con22*	ck351539	1	24	475	yes	yes	yes	6.3	8.5	pancreatic trypsin inhibitor (M)
Hg-Con23	ck349627	1	6	1395	no	no	NA	NA	NA	RNA helicase ATPase (U&M)
Hg-Con24*	co036609	1	3	465	yes	yes	yes	11.1	8.3	none
Hg-Con25*	cb374317	1	5	469	no	yes	yes	8.1	4	glycine-rich protein, *C. elegans* (U)
Hg-Con26*	cb279822	1	1	390	no	yes	yes	8.0	9.9	pancreatic trypsin inhibitor (M)
Hg-Con27	cb378289	1	1	485	no	no	NA	NA	NA	tyrosinase-like *C. elegans*
Hg-Con28*	ca939711	1	14	781	no	yes	yes	18.3	4.3	cellulose binding prot, *H. glycines* (U)
Hg-Sin29	co036788	1	0	367	no	no	NA	NA	NA	sorting nexin (U)
Hg-AK	ay191835	0	257							arginine kinase 1
Hg-cell1	af006052	0	33							endo 1-4 beta glucanase 1
Hg-EF1b	bf014332	0	16							elongation factor 1b
Hg-SYV46	af273728	0	87							CLAVATA-3
Gm-EF1a	ay651886	0	NA							elongation factor 1a
Gm-Actin	bm525891	0	NA							actin

[1] Asterisk indicates a sequence that encodes a peptide with a putative signal peptide and a mature protein <20 kDa.
[2] First six accession numbers are for confirmed sequences for full-length cDNA; all others are representative accession numbers of an *H. glycines* EST.
[3] Number of highly similar clones in the 384 cDNA that were sequenced from the late subtraction library.
[4] Number of highly similar EST in the *H. glycines* database in August 2004.
[5] Length of full-length sequence or consensus sequence minus the poly(A) tail.
[6] Presence of poly(A) tail of greater than 12 bp.
[7] Full-length open reading frame beginning with an AUG and ending with a stop codon.
[8] Presence of a signal peptide predicted by the Accelrys GCG program SPScan.
[9] Molecular size of peptide minus signal peptide where appropriate.
[10] Isoelectric point of mature peptide.
[11] Cutoff for BLAST searches of GenBank uniprot database was $e < 10^{-10}$ with filtering off.
[12] Cutoff for hidden Markov model Pfam library was $e < 10^{-3}$.
[13] NA = Not Applicable.
[14] rRNA were identified by a nearly perfect match with rRNA in the GenBank nt database.

dpi the feeding nematodes have clearly begun to swell and grow. At 20 dpi the nematodes have grown enough to protrude from the root, and by 28 to 30 days the nematodes have completed a life cycle. We initially selected to study the six SCN transcripts that included the greatest number of clones in the late subtraction library (i.e., Hg-Con 1-6). Analysis of the primary sequence for these unigenes revealed that all six included ORFs with predicted amino-terminus signal peptides and five out of six encoded relatively small proteins <15 kDa (Table 1). Further study determined that eight more of the SCN unigenes encoded small proteins (<20 kDa) with putative amino-terminus signal peptides. RNA blot results confirmed the predicted small size of the 13 mRNA and their high expression in later stages of SCN development and not in eggs or J2 (Fig. 1). These 13 small transcripts accounted for 80% of the SCN clones that were sequenced in the late subtraction library. It's

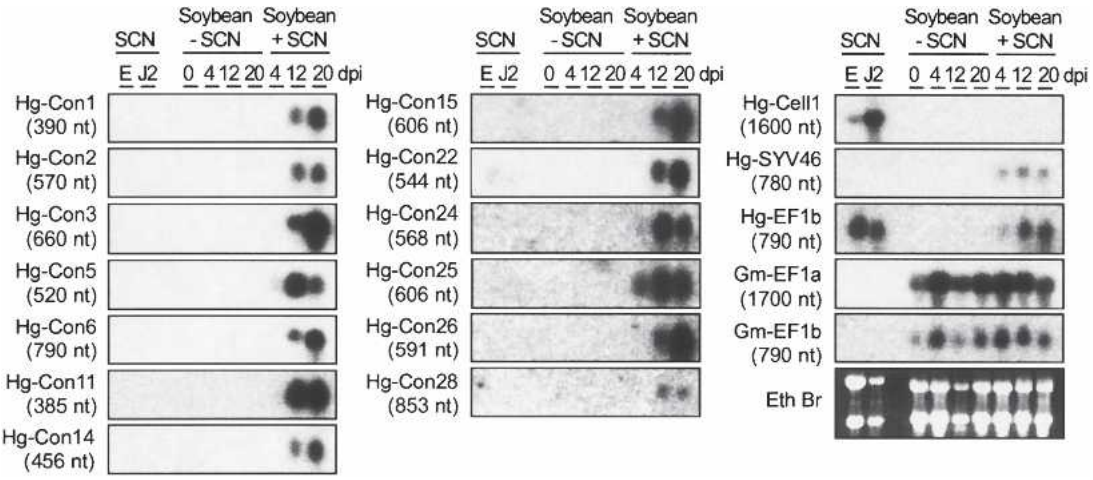

FIG. 1. Northern blot hybridizations for the 13 small SCN transcripts identified in the late subtraction library and five additional control SCN and soybean transcripts. The Hg-Cell1 and Hg-SYV46 probes were for transcripts that have been demonstrated to be abundantly expressed in the esophageal glands of SCN (Gao et al., 2003). The nematode elongation factor 1b (Hg-EF1b) and the soybean elongation factors 1a and 1b were included as constitutive controls that differ in size and expression level. Six identical agarose gels were run and the RNA transferred to six membranes. Each membrane was reprobed several times. The calculated length of the transcripts in nucleotides is in parentheses. E = eggs; –SCN = non-inoculated roots; + SCN = inoculated roots; dpi = days post-inoculated; Eth Br = ethidium bromide stained.

possible that more SCN clones from the subtraction library fell into this category of small ER synthesized proteins, but there was not enough sequence data to complete the ORF. Formation of unigenes for the soybean clones in the late subtraction library did not indicate any particular selection for small mRNA (data not shown). Based on our results, it might be concluded that a large portion of the gene expression associated with later stages of SCN development encode small proteins synthesized on the ER.

The in situ hybridization results and sequence comparisons provide some clues as to the potential function of some of these small peptides. We performed in situ hybridization for the first five small transcripts listed in Table 1 (asterisks). Hg-Con1 did not highlight any particular organ system, and it was not obvious that the antisense probe produced a stronger signal than the sense probe (results not shown). The Hg-Con2 in situ hybridization produced a detectable signal in a location in the anterior portion of a young nematode that might be a dorsal esophageal gland (Fig. 2A at arrow). However, the greatest amount of expression of Hg-Con2 is in a different organ system that appears to be part of the reproductive system (Eisenback, 1985; Handoo, pers. comm.). Association with the reproductive system is further indicated by the observation that the stained structure encloses developing eggs (Fig. 2A,B). Hg-Con3 and Hg-Con6 expression is also associated with the reproductive system (Fig. 2). The organ system associated with the expression of Hg-Con5 begins near the mouth of the nematode and appears to follow a winding path toward the back of the nematode where staining is more diffuse (Fig. 2G). The stained organ system associated with Hg-Con5 did not appear to enclose any eggs (Fig. 2G). It's possible that Hg-Con5 is associated with the digestive track, but this needs to be confirmed with the co-localization of a better-characterized marker. Nevertheless, none of the four in situ hybridizations indicate an expression pattern for a protein that is secreted into the host through the stylet or amphid gland as is SYV46 (Davis et al., 2004; Fig. 2).

BLAST searches with predicted peptides for Hg-Con1, Hg-Con2, Hg-Con3, Hg-Con5, and Hg-Con6 of the uniprot database did not produce E-scores of less than 10^{-10} when filtering out of low-complexity regions was allowed; however, because many of these sequences are rich in one or more amino acids, the searches were repeated with low-complexity filtering turned off. With filtering off, only Hg-Con6 resulted in a significant score. The best match for Hg-Con6 was with an eggshell protein in the fluke (flatworm), *S. mansoni* (Table 1). Also of potential interest was a match with an E score of 2×10^{-9} between Hg-Con6 and a glycine-rich neuropeptide-like protein (*npl-31*) from *C. elegans*. The npl-31 protein is one of a group of neuropeptide-like proteins that was identified in a pattern search for neuropeptides in *C. elegans* (Nathoo et al., 2001). However, Nathoo et al. (2001) noted that the hypodermal expression of this family of *npl* genes was different than other confirmed neuropeptides and that these genes may not encode neuropeptides. Moreover, Hg-Con6 doesn't have the typical KR proteolytic cleavage sites found in most neuropeptide precursor proteins (Nathoo et al., 2001). Neuropeptides tend to be highly conserved in nematodes and other organisms, and most are expressed throughout the organism life cycle (Li et al., 1999a; Nathoo et al., 2001). Based on these criteria, it seems unlikely that any of the 13-selected mRNA encode neuropeptides.

Anti-microbial peptides (AMP), on the other hand,

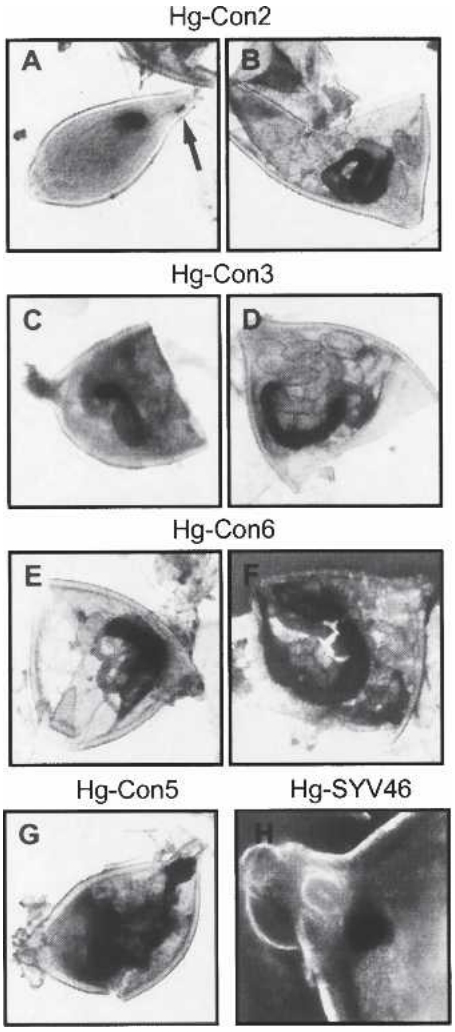

FIG. 2. In-situ hybridization for Hg-Con2, Hg-Con3, Hg-Con5, Hg-Con6, and Hg-SYV46. Nematodes were collected between 15 and 20 d post inoculation of soybean roots growing on Nobel agar with Gamborg's B5 minimal media plus organics and 2% sucrose. The arrow in Figure A points to a minor hybridization signal in the anterior portion of an immature female.

are small proteins and highly divergent except for some very basic structural characteristics (Bulet et al., 2004). AMP are typically less than 25–30 kDa and are often cationic (positively charged) (Bulet et al., 2004). AMP can be classified into three major groups: (i) peptides with an alpha helix conformation (e.g., cecropins), (ii) cyclic and open-ended cyclic peptides with pairs of cysteine residues (e.g., defensins), and (iii) peptides with an over-representation of one or more amino acids (e.g., proline, histidine, or glycine-rich) (Bulet et al., 2004). A cysteine-rich group of AMP (ASABF) and a cecropin-type AMP have been previously identified in nematodes (Andersson et al., 2003; Kato and Komatsu, 1996). Several of the 13 contigs possess characteristics described for AMP (Table 1) (Fig. 3). Hg-Con3, 5, 6, 11, 15, and 26 are strongly cationic with predicted isoelectric points of greater than 9.0 (Table 1). Hg-Con2, 22, 24, and 25 have four or more cysteines that could form disulfide bonds, and Hg-Con1, 3, 5, 6, 11, 15, 24, and 25 are rich in one or more amino acids (Fig. 3). Although several of these have characteristics similar to AMP, anti-microbial activity needs to be determined for each of these before assigning them a function.

Hg-Con22 and Hg-Con26 are most similar to pancreatic (Kunitz) trypsin inhibitors (Table 1) and may serve a function in the digestive track (Marchbank et al., 1998) or secreted outside the nematode to inhibit external serine proteases in the host (Brandon and Friedman, 2002). A BLAST search of the uniprot database with Hg-Con28 identified similarity ($E = 9 \times 10^{-14}$) with an *H. glycines* cellulose binding protein (Gao et al., 2004) that is synthesized in a subventral esophageal gland and, thereby, potentially secreted into the host (Gao et al., 2003). It remains to be determined if the two putative Kunitz trypsin inhibitor-like proteins, the cellulose binding-like protein, or any of the other remaining 6 selected unigenes are synthesized in an esophageal gland or amphid where they might be secreted into the host.

In conclusion, it's interesting that of the 18 complete ORF identified in the 28 contigs from the late subtraction library, 16 include putative signal peptides and 13 of those are <20 kDa (Table 1). The sequence and location of synthesis for three of the more abundant transcripts (Hg-Con2, 3, and 6) suggest a potential role for these proteins in egg development, possibly as AMP that protect the eggs from bacterial or fungal predation.

```
Hg-Con1
  MSPLRFFVALFLITLAILAGTHGTVENSVDGGPNMASPIIRERRWGHGGGGHWG
Hg-Con2
  MVSKLLPVILCCLFLASFALCSPKAADVLKFKREAAISSCGAGQFEYGGGCCTTCTSDADCTGGKLCVPIL
  NLLSVSLCFAPCGN
Hg-Con3
  MKFAPLALLVILLVSLVAAVSANSNNPFDADAADNGGMKRVKRQQQQWRWSPQQQRWVPINIANVNIAQNG
Hg-Con5
  MSPLRFIIAFLLVVLCAFSALITAQEERVDNEPAALRARRGMYGSPYGGGMGGGGPWGGNFGKK
Hg-Con6
  MKLSLFLLALFGIVLVFVAAGDYGYGGYGKKKYGSGSHEYGYGKKKHYKKKHYDSGSYEYGKGSYGKKKYG
  DYGGKKYDDYGGYGKDYGYGGYGKKYSYGSGSHEYGKSYGGYGKSYGGYGKSYSAYGKAYDDYYKKY
Hg-Con11
  MRFFFLLTIALLSVVLLLSPSALAKKRARRSMPVNAAMPEWKILSAEVENDGTA
Hg-Con14
  MFHLRFVVLLLLVIVGLFIGFNFVSADWDKKVAGWDDKKVNKGWEEAEAPLNIRARRGGWDERILGGREQ
Hg-Con15
  MNPTYPMVFFLLVLVAVVLSSVHLKEDKIDVPQRVKRGEYGGGAYGGDGGGVGAPAPGGGGAANIANVXIN
  QQNGCNGRGPQYQWAYTRKTGWVCINIINVIINQQSGGK
Hg-Con22
  MHSXXNFKLFASXFLVLCVVSFVVCQDKPAEACNHPKDXGTGNLALTRWWWDGRXCKQFTYKGQGGNTNNF
  PSKKLCDDVCKP
Hg-Con24
  MKFTLLSILFLALVHIGISGKYDKSDEHDGKYGARPPASGGYHHDNNNYGGGNHYGGGRNHYGGGGSSCDK
  YRPTEHQSTYWSSCGSLLATARSLGCRCGTYKTSETCGFLYLGRRYLCYC
Hg-Con25
  MRTSLLIPLSLFSLITLLFIVSFSEADIEVDSLMPRTKRQCCGGCGYSGGYGGYGNDGGGQCCDACGQCGS
  GGYGSGAIIITTGGGGYGYGGYGDGYCYSCGGGGCYGCCG
Hg-Con26
  MQLSXLSXAFVALFLTSALIEXTEANAPKQNXGTDTNAEKEENVCLLKKHEGSGNGHIRRWWHDKADGECK
  APFYKGRGGNGNRFVSKALCERRCNK
Hg-Con28
  MTKFVCVLLLLFAWTLQIVHPSVVPSDQIYIIESFNRDNSNNQQDNDLKSANGVEDSKMTNGVISQDSAQVV
  TTNGTTELETDPLTVSVELQRGGRAVGNQFNINFTDISYVLVCSARFRISLPSDTTLRSHSKMTAVPNTSD
  QFTLPNGVHLYPGLSHTAEVILSGNGQPEVTILDTLTVLTTEKCPNDV
```

FIG. 3. Open reading frames for the 13 selected sequences described in Table 1 that include amino-terminus signal peptides and mature proteins of less than 20 kDa. The first five sequences were confirmed by 5′ and 3′ RACE, and the other eight are for consensus sequences from the alignment of subtraction library cDNA and EST from the *H. glycines* database (August, 2004). X indicates ambiguities in the consensus sequences. Predicted amino-terminus signal sequences are indicated with a single underline. Cysteine residues in the mature peptide are double underlined.

Literature Cited

Andersson, M., A. Boman, and H. G. Boman. 2003. Ascaris nematodes from pig and human make three antibacterial peptides: Isolation of cecropin P1 and two ASABF peptides. Cellular and Molecular Life Sciences 60:599–606.

Bird, D. McK. 2004. Signaling between nematodes and plants. Current Opinion in Plant Biology 7:372–376.

Boman, H. G. 2003. Antibacterial peptides: Basic facts and emerging concepts. Journal of Internal Medicine 254:197–215.

Brandon, D. L., and M. Friedman. 2002. Immunoassays of soy proteins. Journal of Agricultural and Food Chemistry 50:6635–6642.

Brownlee, D., L. Holden-Dye, and R. Walker. 2000. The range and biological activity of FMRFamide-related peptides and classical neurotransmitters in nematodes. Advances in Parasitology 45:109–180.

Bulet, P., R. Stocklin, and L. Menin. 2004. Anti-microbial peptides: From invertebrates to vertebrates. Immunological Reviews 198:169–184.

Byrd, D. W., T. Kirkpatrick, and K. R. Barker. 1983. An improved technique for clearing and staining plant tissue for detection of nematodes. Journal of Nematology 15:142–143.

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

de Boer, J. M., Y. Yan, G. Smant, E. L. Davis, and T. J. Baum. 1998. In-situ hybridization to messenger RNA in *Heterodera glycines*. Journal of Nematology 30:309–312.

Duguid, J. R., and M. C. Dinauer. 1990. Library subtraction of in vitro cDNA libraries to identify differentially expressed genes in scrapie infection. Nucleic Acids Research 18:2789–2792.

Eisenback, J. D. 1985. Detailed morphology and anatomy of second-stage juveniles, males, and females of the genus *Meloidogyne* (root-knot nematodes). Pp 47–77 *in* J. N. Sasser and C. C. Carter, eds. An advance treatise on *Meloidogyne,* vol. 1. Biology and Control. Raleigh, NC: North Carolina State University Graphics.

Gao, B., R. Allen, E. L. Davis, T. J. Baum, and R. S. Hussey. 2004. Molecular characterisation and developmental expression of a cellulose-binding protein gene in the soybean cyst nematode *Heterodera glycines*. International Journal of Parasitology 34:1377–1383.

Gao, B., R. Allen, T. Maier, E. L. Davis, T. J. Baum, and R. S. Hussey. 2003. The parasitome of the phytonematode *Heterodera glycines*. Molecular Plant Microbe Interactions 16:720–726.

Kato, Y., and S. Komatsu. 1996. ASABF, a novel cysteine-rich antibacterial peptide isolated from the nematode *Ascaris suum*. Purification, primary structure, and molecular cloning of cDNA. Journal of Biological Chemistry 271:30493–30498.

Koehler, S. M., G. L. Matters, P. Nath, E. C. Kemmerer, and M. L. Tucker. 1996. The gene promoter for a bean abscission cellulase is ethylene-induced in transgenic tomato and shows high sequence conservation with a soybean abscission cellulase. Plant Molecular Biology 31:595–606.

Li, C., K. Kim, and L. S. Nelson. 1999a. FMRFamide-related neuropeptide gene family in *Caenorhabditis elegans*. Brain Research 848: 26–34.

Li, C., L. S. Nelson, K. Kim, A. Nathoo, and A. C. Hart. 1999b. Neuropeptide gene families in the nematode *Caenorhabditis elegans*. Annals of the New York Academy of Sciences 897:239–252.

Marchbank, T., T. C. Freeman, and R. J. Playford. 1998. Human pancreatic secretory trypsin inhibitor. Distribution, actions, and possible role in mucosal integrity and repair. Digestion 59:167–74.

Matthews, B. F., M. H. MacDonald, V. K. Thai, and M. L. Tucker. 2003. Molecular characterization of arginine kinase in the soybean cyst nematode (*Heterodera glycines*). Journal of Nematology 35:252–258.

Nathoo, A. N., R. A. Moeller, B. A., Westlund, and A. C. Hart. 2001. Identification of neuropeptide-like protein gene families in *Caenorhabditis elegans* and other species. Proceedings of the National Academy of Sciences 98:14000–14005.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: A laboratory manual, 2nd ed. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

von Heijne, G. 1983. Patterns of amino acids near signal-sequence cleavage sites. European Journal of Biochemistry 133:17–21.

Williamson, V. M., and C. A. Gleason. 2003. Plant-nematode interactions. Current Opinion in Plant Biology 6:327–333.