Journal of Nematology 25(4):548–554. 1993.
© The Society of Nematologists 1993.

Major Sperm Protein Genes from *Globodera rostochiensis*[1]

CHARLES E. NOVITSKI,[2] SHEILA BROWN,[3] RU CHEN,[2] ADAM S. CORNER,[3]
HOWARD J. ATKINSON,[3] AND MICHAEL J. MCPHERSON[3]

Abstract: Three genes in the major sperm protein (MSP) gene family from the potato cyst nematode *Globodera rostochiensis* were cloned and sequenced. In contrast to the absence of introns in *Caenorhabditis elegans* MSP genes, these genes in *G. rostochiensis* contained a 57 nucleotide intron, with normal exon–intron boundaries, in the same relative location as the intron in *Onchocerca volvulus.* The MSP genes of *G. rostochiensis* had putative CAAT, TATA, and polyadenylation signals. The predicted *G. rostochiensis* MSP gene product is 126 amino acids long, one residue shorter than the products in the other species. The comparison of MSP amino acid sequences from four diverse nematode species suggests that *O. volvulus, Ascaris suum,* and *C. elegans* may be more closely related to each other than they are to *G. rostochiensis.*

Key words: *Ascaris suum, Caenorhabditis elegans,* DNA, gene, *Globodera rostochiensis,* major sperm protein, nematode, *Onchocerca volvulus.*

Major sperm protein (MSP) is the principal protein of spermatozoa from the pig roundworm *Ascaris suum* and the microbivorous nematode *Caenorhabditis elegans,* comprising more than 15% of the total cell protein (2,5). MSP is developmentally regulated and has been found only in the testes of these nematodes. In late primary spermatocytes, the 15-kDa MSP protein assembles into filamentous fibrous bodies, within which it is transported to the developing spermatid (11,20). During spermatogenesis, MSP reassembles into 2–3 nm filaments in the pseudopod of the ameboid, nonflagellated nematode spermatozoa (17). MSP has been proposed to have a key role in nematode sperm motility, in view of the absence of significant amounts of actin, myosin, or tubulin in the sperm cells of *A. suum* and *C. elegans* (10,12).

Genes for MSP have been cloned and characterized in *A. suum* (1), the filarial nematode *Onchocerca volvulus* (16), and *C. elegans* (2,4,6). The nucleotide sequence data indicate a polypeptide product of 127 amino acids strongly conserved among these nematode species. The MSP gene family ranges in size from one to three genes in *A. suum* to more than 30 in *C. elegans.* Introns have been found in the MSP genes of *O. volvulus* but not in those of *C. elegans.* The present study elucidates MSP gene structure from a representative of a phytoparasitic nematode lineage, the potato cyst nematode *Globodera rostochiensis.* This information contributes to identifying the most conserved, and thus functionally most essential, regions of the MSP gene product and also takes a step in providing DNA sequence data that address phylogenetic issues.

MATERIALS AND METHODS

DNA library construction: Globodera rostochiensis pathotype Ro1 (kindly supplied by Rothamsted Experimental Station, Harpenden, UK) originally isolated from a field site in the United Kingdom was propagated on *Solanum tuberosum* spp. *tuberosom* cv. Pentland Crown. Hatched juveniles (J2) were obtained from cysts by standard methods (18). DNA was extracted from J2 by proteinase K digestion and phenol extractions (13). The DNA was partially digested with Eco RI, size fractionated by agarose gel electrophoresis, and electroeluted. Fragment sizes 9–23 kb were ligated into λEMBL4 using the manufacturer's protocols (Stratagene, La Jolla, CA).

Library screening and nucleotide sequence de-

Received for publication 17 August 1993.
[1] Financial support was received from the AFRC (grant PG24/238 to HJA and MJM), the USDA NRI Competitive Grants Program (grant 92-37311-8199 to CEN), and the Research Excellence Fund, State of Michigan (grant to CEN).
[2] Department of Biology, Central Michigan University, Mt. Pleasant, MI 48859.
[3] Centre for Plant Biochemistry and Biotechnology, University of Leeds, Leeds LS2 9JT, UK.
We thank Sam Ward (University of Arizona, Tucson) for the gift of an MSP gene probe from *Ascaris.*

termination: An Eco RI fragment containing most of the *Ascaris* MSP cDNA sequence (19) was ^{32}P-labeled using random hexamer primers (Pharmacia, Piscataway, NJ), and this fragment was used to screen the λEMBL4 genomic library using standard procedures (13). Positive plaques were selected from 30,000 plaques screened with the *Ascaris* probe in 0.5 × SSC containing 0.1% SDS at 55 °C, purified via two rounds of further screening, and partially characterized using the restriction enzymes Eco RI, Bam HI, Hind III, and Sma I and Southern analysis (13). Restriction fragments of three clones showing dissimilar restriction patterns (Fig. 1) were subcloned from λEMBL4 into pBluescript (Stratagene) for sequencing. Nucleotide sequence analysis was performed with the dideoxynucleotide method (14), using Sequenase (United States Biochemical, Cleveland, OH) and fractionated on gradient sequencing gels (13). On the basis of good C-terminal amino sequence homology (pro-ile-glu-tyr-asn-(pro, leu) at the C-terminus) between *Caenorhabditis* and *Ascaris*, sequencing was initiated for pBluescript clones carrying *G. rostochiensis* inserts using the highly degenerate primer 5′-CTAT(AG)G(AG)TT-(AG)TA(CT)TCIAT(AGCT)GG-3′. Subsequently, additional sequence-specific primers were synthesized based on the sequence information obtained. These additional primers were CTACATTTT-TCATTGG, GGTTTACAAAAT-TATCTG, CCGGGCGATGACATG, CGTGGCCATCTCCTGC, GATCAGGT-GTGTTGCACC, CTGACTTACCCG-CACG, ATGGCGCAACTTCCTCCAG, CGAAACGCTCCAGCCG, CATTTCG-GCTGGAGCG, GTGGATAATTTG-GAAAAT, and TAGACGAAGATGCGC-CGC. Sequence data were analyzed with the GCG sequence analysis software package (Genetics Computer, Madison, WI).

Results and Discussion

Analysis of genomic DNA libraries: When DNA libraries were constructed for *G. rostochiensis* Ro1 in λEMBL4 and hybridized with the *Ascaris* MSP cDNA sequence, restriction digests and Southern analysis of six positive clones suggested five nonidentical MSP genomic clones. This number of distinct clones was consistent with the 5–6 bands obtained by Southern hybridization analysis of an Eco RI digest of *G. rostochiensis* genomic DNA. The five MSP genes in *G. rostochiensis* fall within the range of the number of MSP gene family members in other nematode species. *Ascaris suum* has one to three copies of the MSP gene, with a single gene characterized (1), two MSP gene family members have been characterized in *Onchocerca volvulus* (16), and more than 30 genes and pseudogenes exist in

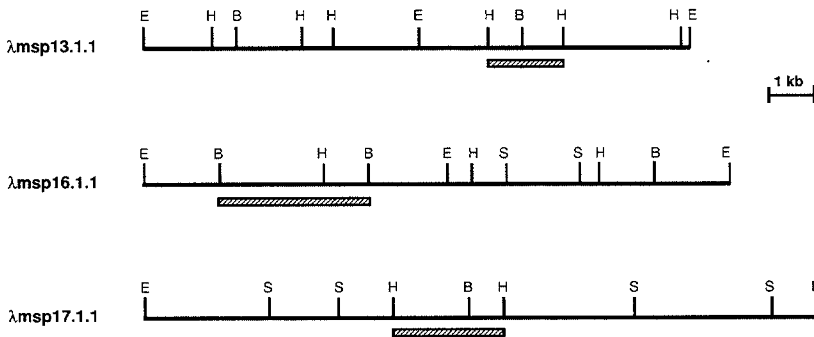

Fig. 1. Restriction maps of three λEMBL4 clones, λmsp13.1.1, λmsp16.1.1, and λmsp17.1.1, containing major sperm protein (MSP) genes of *Globodera rostochiensis*. The sizes of the cloned inserts were 14.5 kb in λmsp13.1.1, 15.7 kb in λmsp16.1.1, and 17.7 kb in λmsp17.1.1. Restriction sites are B, Bam HI; E, Eco RI; H, Hind III; and S, Sma I. Hatched bars under each restriction map indicate the restriction fragments that hybridize to *Ascaris suum* MSP cDNA.

Caenorhabditis elegans (15). On the basis of Southern analysis, the soybean cyst nematode, *Heterodera glycines*, has an estimated five members of the MSP gene family (15). There is a possibility of a low level of contamination of *G. pallida* in a United Kingdom field population of *G. rostochiensis*, which could cause the apparent number of members in the gene family of the species to be overestimated.

Comparison of G. rostochiensis MSP gene sequences: The λEMBL4 clones that gave distinct restriction patterns when hybridized to the *Ascaris* MSP probe (Fig. 1) were subcloned into pBluescript and were subjected to sequence analysis. Sequence information suggested that all three subclones contained MSP genes. These were designated *msp-1* (from λmsp13.1.1), *msp-2* (from λmsp16.1.1), and *msp-3* (from λmsp17.1.1) (Fig. 1). Some, but not all, of the primers supported sequencing reactions of all of the *G. rostochiensis* MSP genes. Each of these genes contained 96- and 282-base coding regions in the first and second exons, respectively, interrupted by a single intron of 57 nucleotides located precisely between the codons for amino acid residues 32 and 33 (Fig. 2). Intron splice junctions contained the expected invariant GT at the donor site and AT at the acceptor site. *The G. rostochiensis* exon–intron boundaries, CGG/GTAAGT(T,C) . . . $(T,A)_n$ . . . $(T,C)_n$ . . . TTTAG/, were nearly identical to the general consensus *C. elegans* boundaries, (A,G)AG/GTAAGTT . . . $(T,A)_n$ . . . TTTCAG/ (3) differing primarily upstream of GT and in an additional run of thymines or cytosines within the *G. rostochiensis* intron. The intron in this *G. rostochiensis* gene differed in only one location from the general consensus for all organisms, (C,A)AG/GT(A,G)AGT . . . $(T,C)_n$ . . . (C,T)AG/ (9). Within the intron, *msp-2* differed from *msp-1* at three nucleotides, *msp-3* differed from *msp-1* at five nucleotides, and *msp-2* and *msp-3* differed at six nucleotides (Fig. 1). The intron sequence was less conserved than the coding region. the intron of *O. volvulus*, which is 158 nucleotides long, is positioned precisely between codons for amino acid residues 33 and 34 (16). The MSP genes of *C. elegans* reportedly do not have introns (5,6).

The upstream region in *msp-1* had putative TATA (at *msp-1* nucleotide −104, with numbering relative to the first base in the ATG translation initiation site as +1) and CAAT (at *msp-1* nucleotide −205) regulatory sequences (Fig. 3) and the sequence AACAATCTACA (at −44) 72% homologous to the *C. elegans* initiation of transcription conserved sequence CATAATCTTCA (4). Therefore, the adenine at −34 may serve as the transcription initiation site. The putative TATA box is further upstream (70 bp) of the transcription initiation site than the 31 nucleotides upstream in *C. elegans* MSP genes (4). The putative CAAT box was not reported for *C. elegans* (4), but a comparison with the data for *G. rostochiensis* suggests that insufficient sequence upstream of the coding region in *C. elegans* may have been obtained to observe this regulatory signal. A notable region of poor conservation among the three *Globodera* MSP genes was that between the putative CAAT and TATA sequences, but the significance of this observation remains to be determined. This region may serve no function, or there may be gene-specific differences in binding to different gene regulatory factors. There was also a putative AATAAAA polyadenylation signal sequence for *msp-1* 154 nucleotides downstream of the coding region. The sequence around the translation start codon, CAACAATGGC, was identical in all three genes. This sequence matches well the consensus ribosome binding site for initiation, with 80% identity and a purine in the −3 position (7).

The predicted length of the polypeptide product of each of the *G. rostochiensis* MSP genes is 126 amino acid residues, and nucleotide sequence alignments clearly indicate that these genes are highly conserved within the coding region (Fig. 2). Gene *msp-1* differs from *msp-2* by five nucleotides, leading to four amino acid residue differences, glu_7 (*msp-1*) instead of gly_7

```
                .         .         .         .         .         .
msp-1                          <cgatgacatgtaggcgctctgggcgccgccagctgc  -401

                .         .         .         .         .         .
msp-1  tggagctgctccgccgggcgatgacatgtaggcgctctgggcgccgccagctgctggagc  -341

                .         .         .         .         .         .
msp-1  tgctccgccgggcgatgacatgtaggcgctctgatcgctcatgtttggtgaattgctgcg  -281
msp-3                <*****************c**********************t****

                .         .         .         .         .         .
msp-1  acgaactgcgtagacgagcttgctgtggactgaagttgttgctgtgttgcttctgaagat  -221
msp-2                   <tgctg*tg*acg****************************gc
msp-3  g*****g*****************************************************

             CAAT         .         .         .         .         .
msp-1  tgg---caatggt--gatgaatgag-caaatccacttccttatatacgcttttcaataat  -161
msp-2  ***aga*******--*c****ac*c-********t******t*****a*******c*g**
msp-3  ***---*******gt**c******c-********t******t*****at******c*g**

                .         .         .         .         .     TATA.
msp-1  gaaattcgtttcaaaaaaaataaaga----gtgac-aaacgctccagccgaaatgtatat  -101
msp-2  **************************aatt**t**g************************
msp-3  *******t**a******tt*ag**---att*****--*****--**-----**a******

                .         .         .         .         . site for.
msp-1  aagcgaatggattttccaaattatccaca-tcagcaatcaacaaatcaacgttctcaaca  -41
msp-2  *****************************-******************************
msp-3  *******c*********************t*******a**********************

                                               MetAlaGlnLeuProProGl
      transcription initiation    .         .         .         .
msp-1  atctacacatcaagcgagctcgacaacagcatcaacaacaATGGCGCAACTTCCTCCAGA  +20
msp-2  ***********************************************************G
msp-3  ************************************************************

       uAspIleAlaThrMetProAlaGlnLysValValPheAsnAlaProPheAspAsnLysAl
                .         .         .         .         .         .
msp-1  AGACATCGCAACAATGCCCGCCCAGAAGGTCGTCTTCAACGCACCGTTCGACAACAAGGC  +80
msp-2  *******************AA***************************************
msp-3  ************************************************************

       aThrTyrTyrValArg
                .      GT .         .         .         .         .
msp-1  CACCTACTACGTGCGGgtaagtcagcatcgtttgaccggggaacataatctcatttaatc  +140
msp-2  **********************t*************************************
msp-3  **********************t***********c*g***********************

                    IleIleAsnProGlyThrLysArgIleGlyPheAlaPheLysThrTh
                . AG      .         .         .         .         .
msp-1  ccttctaatttagATCATCAATCCTGGGACAAAGCGCATCGGCTTCGCGTTCAAAACGAC  +200
msp-2  ****tc*******G*******************T**************************
msp-3  *t*c********************************************************

       rLysProLysArgIleAsnMetAsnProProAsnGlyValLeuGlyProLysGluSerVa
                .         .         .         .         .         .
msp-1  CAAACCGAAGCGCATCAACATGAACCCTCCGAATGGCGTGCTCGGCCCCAAGGAGTCCGT  +260
msp-2  ************************************************************
msp-3  ************************************************************

       lAsnValAlaIleSerCysAspAlaPheAspProSerSerGluAspThrLysGlyAspAr
                .         .         .         .         .         .
msp-1  CAACGTGGCCATCTCCTGCGACGCGTTCGACCCCAGCTCCGAGGACACCAAAGGCGACCG  +320
msp-2  ************************************************************
msp-3  **********************************************T*************

       gValThrValGluTrpCysAsnThrProAspProAlaAlaAlaAlaPheLysLeuGluTr
                .         .         .         .         .         .
msp-1  TGTGACCGTGGAATGGTGCAACACACCTGATCCGGCGGCTGCCGCGTTCAAGCTCGAGTG  +380
msp-2  ************************************************************
msp-3  ************************************************************

       pPheGlnGlyAspGlyMetValArgArgLysAsnLeuProIleGluTyrAsnValEnd
                .         .         .         .         .         .
msp-1  GTTCCAGGGAGACGGTATGGTGCGCCGCAAGAACTTGCCGATCGAGTACAACGTCTAGac  +440
msp-2  ************************************************************
msp-3  ************************************************************

                .         .         .         .         .         .
msp-1  gaagatgcgccgccgcattacgaccagatgcagaagataggatccgcccccccctccaat  +500
msp-2  *************----a*****************t****ataggat**at***c*g*t*
msp-3  *********************************************-**a***********

                .         .         .         .         .         .
msp-1  atgtaaatgccctcatgttgtatccccctgaacagataattttgtaaaccaccatatcat  +560
msp-2  cccgc*tattgtatc*cccc*ga*gtgaaaccatatatcatcatgtatgtttgttcatat
msp-3  ************************************************************

                .         .         . polyA           .         .
msp-1  cttatgtctgttcatatgtatgtaaatgtgtaataaaatgtagctttgattggcgcaaca  +620
msp-2  gtaaatgtgtaaataaaatgttgtggtgaggggcgcgacaaaaaaa>
msp-3  ******c****************-*******************t****************

                .         .         .         .         .         .
msp-1  aaaaaatttgggaaaaatggttgacacgttctgaattttgggccttgctttgtttccca  +680
msp-3  **********at************g***cct*******-*t*****************>

                .         .         .         .         .         .
msp-1  attcatgtccttcccgcaccaatgaaaaatgtagaaatttgttagagtaggggcgggcga  +740

                .         .         .         .         .         .
msp-1  atgtattgaaaacatggtttcaatttttgagaaaatttttgtcataattcctgcaaaaa  +800

                .         .         .         .         .         .
msp-1  caggcccatagcatttagtggcctctttagactacaaaagattgaagccgtacacacaca  +860

                .         .         .         .         .         .
msp-1  cgcgatacaacggattgttctttccccagtta                              +888
```

FIG. 2. Nucleotide sequence of three MSP genes (*msp-1*, *msp-2*, and *msp-3*) of *Globodera rostochiensis*. Conserved putative CAAT box, TATA box, transcription initiation site, intron splice boundaries GT and AG, and polyadenylation signal are indicated. An asterisk indicates a nucleotide identical to the one in *msp-1* at that position. Dashes indicate where insertions were made to improve alignment. Numbering refers to the last nucleotide in each line of the *msp-1* nucleotide sequence and is in reference to the first base A (+1) in the ATG translation initiation site. Nucleotides of the two coding regions in the two exons are in uppercase and other nucleotides are in lowercase. The amino acid sequence of the predicted *msp-1* gene product is given in three-letter code.

```
                        10         20         30         40
G.r. msp-1      MAQ-LPPEDI ATMPAQKVVF NAPFDNKATY YVRIINPGTK
G.r. msp-2      MAQ-LPPGDI ATMPNQKVVF NAPFDNKATY YVRVINPGTN
G.r. msp-3      MAQ-LPPEDI ATMPAQKVVF NAPFDNKATY YVRIINPGTK
C.e. msp-3      MAQSVPPGDI QTQPGTKIVF NAPYDDKHTY HIKVINSSAR
A.s. msp        MAQSVPPGDI NTQPSQKIVF NAPYDDKHTY HIKITNAGGR
O.v. msp-1      MAQSVPPGDI HTQPSSKIVF NAPYDDKHTY HINITNAGGR
O.v. msp-2      MAQSVPPGDI HTQPGSKIVF NAPYDDKHTY HIKITNAGGR
                ***  ** **  * *  * ** *** * * **      *

                        50         60         70         80
G.r. msp-1      RIGFAFKTTK PKRINMNPPN GVLGPKESVN VAISCDAFDP
G.r. msp-2      RIGFAFKTTK PKRINMNPPN GVLGPKESVN VAISCDAFDP
G.r. msp-3      RIGFAFKTTK PKRINMNPPN GVLGPKESVN VAISCDAFDP
C.e. msp-3      RIGYGIKTTN MKRLGVDPPC GVLDPKEAVL LAVSCDAFAF
A.s. msp        RIGWAIKTTN MRRLSVDPPC GVLDPKEKVL MAVSCDTFNA
O.v. msp-1      RIGWAIKTTN MKRLGVDPPC GVLDPNEKVL MAVSCDTFDA
O.v. msp-2      RIGWAIKTTN MKRLGVDPPC GVLDPNEKVL MAVSCDTFDA
                ***   ***    *    **  *** * * *   * *** *

                        90        100        110        120
G.r. msp-1      SSEDTKGDRV TVEWCNTPDP AAAAFKLEWF QGDGMVRRKN
G.r. msp-2      SSEDTKGDRV TVEWCNTPDP AAAAFKLEWF QGDGMVRRKN
G.r. msp-3      SSEDSKGDRV TVEWCNTPDP AAAAFKLEWF QGDGMVRRKN
C.e. msp-3      GQEDTNNDRI TVEWTNTPDG AARQFRREWF QGDGMVRRKN
A.s. msp        ATEDLNNDRI TIEWTNTPDG AAKQFRREWF QGDGMVRRKN
O.v. msp-1      TREDINNDRI TVEWTNTPDG AAKQFRREWE QGDGMVRRKN
O.v. msp-2      TREDINNDRI TIEWTNTPDG AAKQFRREWF EGDGMVRRKN
                  **   **  * ** ****  **  *  **   *********

                    127
G.r. msp-1      LPIEYNV
G.r. msp-2      LPIEYNV
G.r. msp-3      LPIEYNV
C.e. msp-3      LPIEYNP
A.s. msp        LPIEYNL
O.v. msp-1      LPIETNL
O.v. msp-2      LPIETNL
                **** *
```

FIG. 3. A comparison of the predicted amino acid sequences of the products from MSP genes *Globodera rostochiensis msp-1, msp-2,* and *msp-3, Caenorhabditis elegans msp-3, Ascaris suum msp,* and *Onchocerca volvulus msp-1* and *msp-2.* The gene numbering systems in each species are unrelated to each other. A dash indicates where an insertion has been made to improve alignment. An asterisk indicates that the amino acid residues are identical in all seven genes at that position. The five underlined residues indicate positions at which products of *Globodera rostochiensis msp-2* and *msp-3* differ from that of *msp-1.*

(*msp-2*), ala_{14} instead of asn_{14},lys_{39} instead of asn_{39}, and the conservative change ile_{33} instead of val_{33}. Gene *msp-3* differs from *msp-1* by one nucleotide, leading to one conservative amino acid residue difference of thr_{84} instead of ser_{84}.

The nucleotide sequence for the MSP gene coding region is known for a single gene in *Ascaris suum* (1), two genes in *Onchocerca volvulus* (16), and a representative gene from *C. elegans* (*msp-3*) (1). The predicted amino acid sequences produced from these four MSP genes are 127 residues long, in comparison to 126 residues in *G. rostochiensis.* Seventy of the 127 residues (55%) are invariant in these three nematode species, including a conserved 13-residue block at the C-terminal end of the polypeptides (Fig. 3). Conserved amino acids are likely indicators of functionally important regions of polypeptides. In addition, the presence of introns separating codons 32 and 33 in *G. rostochiensis* and codons 33 and 34 in *Onchocerca* may reflect two separate functional groups at the N-terminal and C-terminal ends of the polypeptides. The alignment of the polypeptide sequences of MSP gene product from *Onchocerca* and *Globodera* reveals that the introns are in fact at the same position

within the MSP gene, because of the absence of a serine codon at the fourth codon in *G. rostochiensis*.

The percentage of identical amino acid residues between pairs of MSP polypeptides suggest that the animal-parasitic nematodes *Onchocerca* and *Ascaris* are the most closely related to each other (91%), and both are more closely related to the microbivorous soil nematode *C. elegans* (80–82%) than any of the three species are to the plant-parasitic nematode *G. rostochiensis* (60–61%). These data are consistent with the divergence of the ancestors of Ascaridida from those of Filarioidea, after their divergence from Archi-nemata lines giving rise to Rhabditida and Tylenchida, as envisaged in the classification of Maggenti (8). Notwithstanding the caution required in extrapolation of results based on one gene and based on only one genus in each feeding type, the data also suggest the possibility of the divergence of the ancestors of Rhabditida, Ascaridida, and Filarioidea and after their divergence from the Archi-nemata line giving rise to Tylenchida. Because the MSP introns of *G. rostochiensis* and *O. volvulus* are in precisely the analogous locations within the genes, the absence of introns in *C. elegans* probably results from the loss of the intron in a single event in that lineage, rather than two separate intron insertion events in the ancestors of the other two species. The similarity in the nucleotide sequences of the *G. rostochiensis* MSP gene family members suggest that the genes may have arisen through gene duplication after the divergence of this species from the others mentioned here, followed by neutral genetic drift and possibly selection to generate the sequence differences. In addition, because in each species all MSP gene family members are consistent in either having or lacking an intron, gene duplication appears to have occurred over a shorter time frame than did the divergence of these species. The MSP genes and their protein products have been highly conserved throughout nematode evolution. This conservation probably reflects a key role for MSP in sperm motility common to very different nematode species.

Literature Cited

1. Bennett, K. L., and S. Ward. 1986. Neither a germ line-specific nor several somatically expressed genes are lost or rearranged during embryonic chromatin diminution in the nematode *Ascaris lumbricoides* var. *suum*. Developmental Biology 118:141–147.

2. Burke, D. J., and S. Ward. 1983. Identification of a large multigene family encoding the major sperm protein of *Caenorhabditis elegans*. Journal of Molecular Biology 171:1–29.

3. Emmons, S. W. 1988. The genome. Pp. 47–79 *in* W. B. Wood, ed, The nematode *Caenorhabditis elegans*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

4. Klass, M., R. Ammons, and S. Ward. 1988. Conservation in the 5′ flanking sequence of transcribed members of the *Caenorhabditis* major sperm protein gene family. Journal of Molecular Biology 199:15–22.

5. Klass, M. R., B. Dow, and M. Herndon. 1982. Cell-specific transcriptional regulation of the major sperm protein in *Caenorhabditis elegans*. Developmental Biology 93:152–164.

6. Klass, M. R., S. Kinsley, and L. C. Lopez. 1984. Isolation and characterization of a sperm-specific gene family in the nematode *Caenorhabditis elegans*. Molecular and Cellular Biology 4:529–537.

7. Kozak, M. 1984. Compilation and analysis of sequences upstream from the translation start site in eukaryotic mRNAs. Nucleic Acids Research 12:857–872.

8. Maggenti, A. R. 1991. Nemata: Higher classification. Pp. 147–187 *in* W. R. Nickle, ed. Manual of agricultural nematology. New York: Marcel Dekker.

9. Mount, S. M. 1982. A catalogue of splice junction sequences. Nucleic Acids Research 10:459–472.

10. Nelson, G. A., T. M. Roberts, and S. Ward. 1982. *Caenorhabditis elegans* spermatozoan locomotion: Amoeboid movement with almost no actin. Journal of Cell Biology 92:121–131.

11. Roberts, T. M., F. M. Pavalko, and S. Ward. 1986. Membrane and cytoplasmic proteins are transported in the same organelle complex during nematode spermatogenesis. Journal of Cell Biology 102: 1787–1796.

12. Roberts, T. M., and S. Ward. 1982. Centripetal flow of pseudopodial surface components could propel the amoeboid movement of *Caenorhabditis elegans* spermatozoa. Journal of Cell Biology 92:113–120.

13. Sambrook, J., E. F. Fritch, and T. Maniatis. 1989. Molecular cloning, 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

14. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the USA 74:5463–5467.

15. Scott, A. L., J. Dinman, D. J. Sussman, and S.

Ward. 1989. Major sperm protein and actin genes in free-living and parasitic nematodes. Parasitology 78: 471–478.

16. Scott, A. L., J. Dinman, D. J. Sussman, P. Yenbutr, and S. Ward. 1989. Major sperm protein genes from *Onchocerca.* Molecular and Biochemical Parasitology 36:119–126.

17. Sepsenwol, S., H. Ris, and T. M. Roberts. 1989. A unique cytoskeletal associated with crawling in the ameboid sperm of the nematode *Ascaris suum.* Journal of Cell Biology 108:55–56.

18. Southey, J. F., ed. 1986. Laboratory methods for work with plant and soil nematodes. London: Her Majesty's Stationery Office.

19. Ward, S., D. J. Burke, J. E. Sulston, A. R. Coulson, D. G. Albertson, D. Ammons, M. Klass, and E. Hogan. 1988. Genomic organization of major sperm protein genes and pseudogenes in the nematode *Caenorhabditis.* Journal of Molecular Biology 199:1–13.

20. Ward, S., and M. Klass. 1982. The location of the major sperm protein in *Caenorhabditis elegans* sperm and spermatocytes. Developmental Biology 92:203–208.

GenBank Accession

The GenBank accession numbers for the sequences described in this paper are as follows: *msp-1* (λ13.1.1), L24499 (1,316 bp); *msp-2* (λ16.1.1), L24500 (860 bp), and *msp-3* (λ17.1.1), L24501 (984 bp).