RESEARCH NOTE

Sequencing and Expression Analysis of a *Schistosoma mansoni* Gene Homologue to a *Drosophila* Gene Involved in Germ Plasm Assembly

Élida ML Rabelo/+, Adriano BS Hobaika, Paulo Marcos Z Coelho, Sérgio DJ Pena*

Departamento de Parasitologia *Departamento de Bioquímica e Imunologia, ICB, Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627, 31270-901 Belo Horizonte, MG, Brasil

Key words: *Schistosoma mansoni* - mago nashi - germ plasm - gene characterization - sequencing

Genes coding for proteins involved in gene regulation and/or development are of great interest in the study of the biology of *Schistosoma mansoni*. This trematode is the etiologic agent of schistosomiasis and presents a complex life cycle with drastic morphologic changes between stages. Recently, some strains have become resistant to the drugs currently in use to eradicate the disease (D Cioli et al. 1995 *Pharmac Ther 68:* 35-85). The strategy of gene discovery program in *S. mansoni* by using the EST (expressed sequence tag) approach (GR Franco et al. 1995 *Gene 152*: 141-147) has been very efficient in the discovery of new *S. mansoni* genes, which were unlikely to be identified using classical procedures based on phenotype. A class of genes that interested us particularly were those that in other organisms were known to be involved in the regulation of embryogenesis. Among these, we selected one that presented a high homology to a *Drosophila* gene named mago nashi. In diptera this gene is involved in the process of germ plasm assembly and its mutation results in sterility of F1 progeny and also in the formation of the perpendicular axes (RE Boswell et al. 1991 *Development 113*: 373-384). We reasoned that this gene might conceivably play a role in the morphogenetic changes seen in the life cycle of the parasite. We thus decided to characterize *S. mansoni* mago nashi further by obtaining its full length cDNA and genomic sequences, as well as studying its expression pattern at the different life stages of the worm.

The whole cDNA was sequenced in both directions yielding 485 nucleotides (nt) that coded for a protein of 146 amino acids with 84% of homology to the *Drosophila* homologue (Fig. 1). Procedures for plasmidial DNA preparation, sequencing and analysis of sequences have been previously described (Franco *loc. cit.*). From the cDNA sequence, primers were designed and the genomic gene amplified from *S. mansoni* total DNA. The genomic amplified product was bigger than the correspondent cDNA amplified product when two different pairs of primers were used (Fig. 2). The PCR amplification products were cloned in a plasmid vector (pUC18) by using the Pharmacia Sure clone kit and sequenced in both directions using fluorescent primers in an automated DNA sequencer. Consistent with the results of Fig. 2, three introns were identified in the genomic sequence, two of 34 nt and one with 33 nt producing a total of a 101 nt of intron sequences (Fig. 3).

Our next step was to investigate the gene expression pattern at the different life cycle stages. Through the RT-PCR technique using cDNA obtained from different stages and specific primers, it was shown that mago nashi is expressed in all stages studied i.e. egg, schistosomula and adult worm (data not shown). This was not unexpected although this gene was first identified in drosophila embryos, it was also identified in adult flies (PA Newmark & RE Boswell 1994 *Development 120*: 1303-1313). The gene had also been shown to be expressed in a large variety of human adult tissues such as lung, kidney, liver, heart, pancreas, brain and placenta (X- Zhao et al. 1998 *Genomics 47*: 319-322).

Since this work was started, the mago nashi gene has been identified in a variety of different organisms. These include mouse, human, *Caenorhabditis elegans*, *Brugia malayi*, *Arabidopsis thaliana* and *Oryza sativa*. Alignment of the conceptual translations of these mago genes

This investigation received financial support from UNDP/WORLD BANK/WHO Special Programme for Research in Tropical Diseases (ID no. 940325, 940751); Fapemig Process No. CBS 2205/96; CNPq and PADCT. Sequence data from this article have been deposited in dBEST and GenBank under accession no. AAA125719
+Corresponding author. Fax: +55-31-499.2970.
Received 4 May 1998
Accepted 31 August 1998

reveals that 63% of the residues are identical in all homologues identified and many of the remaining positions show only conservative substitutions (DR Micklem et al. 1977 *Current Biology 7*: 468-478).

Although a specific function has not been assigned to mago nashi, the fact that the protein seems highly conserved during evolution between animals and plants and is not detected in yeast or bacteria, suggests that it plays some fundamental role in multicellular eukaryotes. The expression of the protein and antibody production will permit the protein immunolocalization shedding some light about the possible functional role it is playing in *S. mansoni.*

Acknowledgements: to Dr José Miguel Ortega for helping with the sequence alignments.

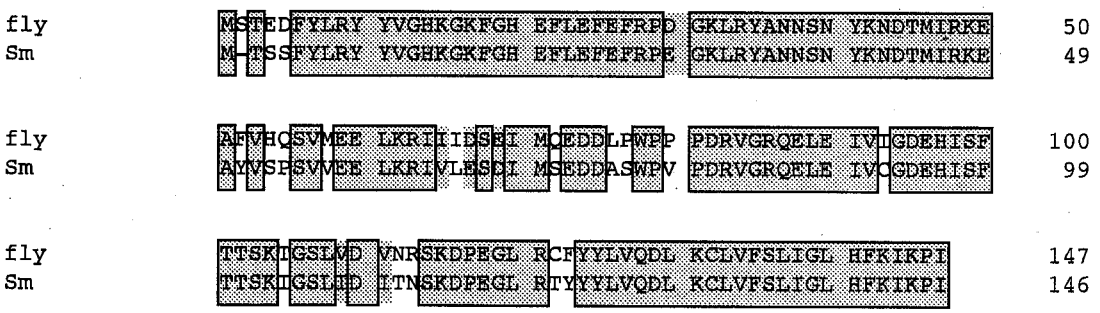

Fig. 1: homology between the mago nashi protein from *Drosophila* and the putative protein codified by the *Schistosoma mansoni* gene. The sequences of both proteins were aligned using the program GeneWorks. The alignment shows 84% of homology.

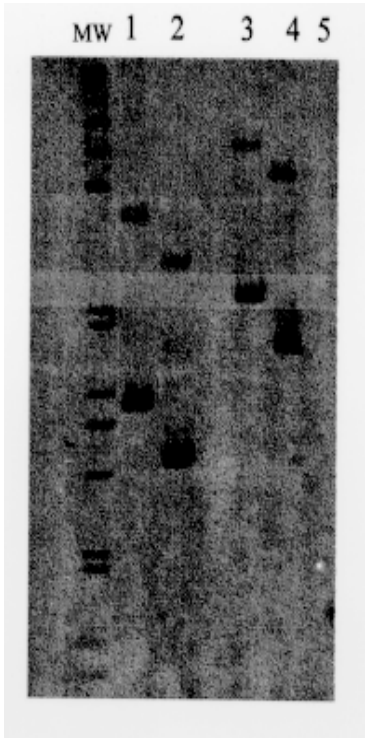

Fig. 2: genomic and cDNA amplification of the *Schistosoma mansoni* mago nashi gene. PCR reactions were carried out using two different pairs of primers based on the cDNA sequence. The pair one gives a product of 300 bp and the pair two gives a product of 450 bp. MW: 1Kb ladder (BRL). Lanes 1 and 2: cDNA amplification using pair one and two respectively; lanes 3 and 4: genomic amplification using pair one and two respectively; lane 5: PCR negative control. The PCR products were resolved on a 6% polyacrylamide gel and silver stained.

```
            1                                                   50
    cDNA    TCGTCCGTCG AGATAATTCA AAAAAAGTGA TTAACAAATG ACAAGCTCAT
  genomic   .......... .......... .....AGTGA TTAACAAATG ACAAGCTCAT
Consensus   .......... .......... .....AGTGA TTAACAAATG ACAAGCTCAT

            51                                                  100
    cDNA    TTTATTTGCG ATACTACGTT GGTCATAAAG GCAAGTTTGG ACATGAATTT
  genomic   TTTATTTGCG ATACTACGTT GGTCATAAAG GCAAGTTTGG ACATGAATTT
Consensus   TTTATTTGCG ATACTACGTT GGTCATAAAG GCAAGTTTGG ACATGAATTT

            101                                                 150
    cDNA    CTTGAGTTCG AGTTCAGACC TGAAGG.... .......... ..........
  genomic   CTTGAGTTCG AGTTCAGACC TGAAGGTAAA GTTTACTCAC TTGTAACAAA
Consensus   CTTGAGTTCG AGTTCAGACC TGAAGG.... .......... ..........

            151                                                 200
    cDNA    .........C AAGTTAAGAT ATGCTAACAA CTCCAATTAT AAAAATGACA
  genomic   TTCTTTAGGC AAGTTAAGAT ATGCTAACAA CTCCAATTAT AAAAATGACA
Consensus   .........C AAGTTAAGAT ATGCTAACAA CTCCAATTAT AAAAATGACA

            201                                                 250
    cDNA    CAATGATTCG CAAAGAGG.. .......... .......... ..........
  genomic   CAATGATTCG CAAAGAGGTA CAATTACTTT AAAATTCAAA TAATGCTCCA
Consensus   CAATGATTCG CAAAGAGG.. .......... .......... ..........

            251                                                 300
    cDNA    ..CGTATGTT AGCCCCTCTG TAGTGGAAGA GCTGAAGAGA ATCGTGTTGG
  genomic   GGCGTATGTT AGCCCCTCTG TAGTGGAAGA GCTGAAGAGA ATCGTGTTGG
Consensus   ..CGTATGTT AGCCCCTCTG TAGTGGAAGA GCTGAAGAGA ATCGTGTTGG

            301                                                 350
    cDNa    AGAGCGATAT TATGTCGGAA GATGATGCTT CATGGCCAGT ACCTGACAGA
  genomic   AGAGCGATAT TATGTCGGAA GATGATGCTT CATGGCCAGT ACCTGACAGA
Consensus   AGAGCGATAT TATGTCGGAA GATGATGCTT CATGGCCAGT ACCTGACAGA

            351                                                 400
    cDNA    GTTGGCCGTC AAGAGCTCGA AATTGTTTGT GGCGATGAAC ACATATCTTT
  genomic   GTTGGCCGTC AAGAGCTCGA AATTGTTTGT GGCGATGAAC ACATATCTTT
Consensus   GTTGGCCGTC AAGAGCTCGA AATTGTTTGT GGCGATGAAC ACATATCTTT

            401                                                 450
    cDNA    CACAACTTCA AAAATAGGAT CCCTGATTGA TATTACGAAT AGCAAG....
  genomic   CACAACTTCA AAAATAGGAT CCCTGATTGA TATTACGAAT AGCAAGTAAG
Consensus   CACAACTTCA AAAATAGGAT CCCTGATTGA TATTACGAAT AGCAAG....

            451                                                 500
    cDNA    .......... .......... .......... GATCCTGAGG GATTAAGAAC
  genomic   TTACTCATTT AATCTAATTT TGAACAAAGG GATCCTGAGG GATTAAGAAC
Consensus   .......... .......... .......... GATCCTGAGG GATTAAGAAC

            501                                                 550
    cDNA    ATATTACTAC TTAGTTCAGG ACCTCAAGTG TCTGGTGTTC TCGCTGATTG
  genomic   ATATTACTAC TTAGTTCAGG ACCTCAAGTG TCTGGTGTTC TCGCTGATTG
Consensus   ATATTACTAC TTAGTTCAGG ACCTCAAGTG TCTGGTGTTC TCGCTGATTG

            551                                   586
    cDNA    GATTGCACTT CAAAATAAAG CCAATATAAT ACGTTC
  genomic   GATTGCACTT CAAAATAAAG CCAATATAAT A.....
Consensus   GATTGCACTT CAAAATAAAG CCAATATAAT A.....
```

Fig. 3: sequence alignment between the *Schistosoma mansoni* mago nashi cDNA and the genomic sequence. The full length cDNA and genomic sequences were aligned using the program "multiple sequence alignment with hierarchical clustering" (F Corpet 1988 *Nucl Acids Res 16:* 10881-10890). The dots represent intron sequences.