Identification of a *mariner*-like repetitive sequence in *C.elegans*

Margaret M.Sedensky, Susan J.Hudson, Barbara Everson and Phil G.Morgan*

Departments of Anesthesiology and Genetics, Case Western Reserve University and University Hospitals, 11100 Euclid Avenue, Cleveland, OH 44106, USA

Received October 19, 1993; Revised and Accepted March 5, 1994

ABSTRACT

A repetitive element in *C.elegans* has been found that bears high homology to the element *mariner* of *Drosophila mauritiana* (EMBL accession number X77804). This element is present in about 20 copies in the N2 strain of *C.elegans*, and appears in roughly equal copy numbers in the related strain B0 and in the hybrid strains RW7097 and TR679. There is only one copy of this MLE in three related species of *Caenorhabditis*. A cDNA of this *mariner*-like element (MLE) codes for a protein with 58% homology to the *Drosophila* transposase. The *mariner*-like element is not mobile in N2. This class of elements has now been described in insects, planaria and nematodes (GenBank accession number M98552 and this report).

INTRODUCTION

Transposable elements are a ubiquitous feature of the genome of both pro- and eukaryotes, and pose many questions concerning their phylogenetic distribution. (3,4,5,6,7) In the nematode *C.elegans*, six transposons, Tc1–Tc6, have been identified. (8,9,10,11,12,13) A survey of these elements within the phylum shows a very patchy distribution of different transposons among different nematodes. (14) This pattern of distribution cannot be explained by strictly vertical transmission. These *C.elegans* transposons also bear a resemblance to those of other phyla. For example, the best characterized transposon of *C.elegans*, Tc1, contains short terminal inverted repeats (15), a feature shared not only by other *C.elegans* transposons (10,13), but by many eukaryotic transposable elements across a wide phylogenetic spectrum. (13,16) Tc1 has a single open reading frame and its terminal repeats share sequence identity with an element of a related nematode species (17), as well as to the *Drosophila* sequence, HB1. (17,18) Novel computer search methods have identified homologies to the putative transposase, TcA, in an array of organisms ranging from bacteria to invertebrates to vertebrates. (7) Whether the Tc1 family of transposons demonstrates a puzzling mix of vertical and horizontal transmission of a very ancient ancestor, or convergent evolution, is not clear.

Patchy distribution of a well-characterized transposable element is also found within the insect world. The *mariner* element was

first discovered by Hartl and co-workers in *Drosophila mauritiana*. (19) It is absent in *D.melanogaster*, but seen extensively within the genus. (20) Lidholm *et al.* have also identified a *mariner*-like element (MLE) in the lepidopteran, *Hyalophora cecropia*. (21) They found the overall DNA homology was 48% between *mariner* elements in *D.mauritiana* and *H.cecropia*. They were unable to determine if the MLE were mobile, but did show that the element was present in about 1000 copies. Robertson performed an extensive survey to determine the extent of *mariner* in insects. (1) Using sequences common to the MLE from *Hyalophora* and *Drosophila* as primers for polymerase chain reactions, the author found equivalent regions in ten other species of varying diversity. The different MLEs were most easily divided into four groups based on conserved DNA sequence. He concluded that both vertical and horizontal transmissions were responsible for the patterns of *mariner* inheritance. A recent report by Garcia-Fernandez *et al.* has identified a *mariner* like element (MLE) in planaria, showing that MLEs are not restricted to insects. (2)

We are attempting to clone and sequence the *unc-80* gene in the nematode *C.elegans*. During this work, we identified sequence homology between a repetitive element in *C.elegans* and *mariner* of *D.mauritiana*. We have investigated its distribution and activity in different strains of *C.elegans* and different species of *Caenorhabditis*. In addition, another MLE has been identified by the *C.elegans* genome sequencing project. (GenBank accession number M98552) The existence of *mariner* in a highly conserved fashion between nematodes and insects raises questions as to its mode of transmission.

METHODS

Nematodes

Cultures of *C.elegans* were grown and maintained as previously described by Brenner. (22) The wild type strain, N2, the closely related Bergerac strain, B0, and a hybrid strain, RW7097, were obtained from the *Caenorhabditis* Genetics Center in Columbia, MO. Unlike N2, B0 possesses a high copy number of Tc1, which are actively mobile. (23) B0 thus has a high forward mutation rate. RW7097 originated by crossing B0 to N2 and selecting for a high frequency of spontaneous mutation in the hybrid

*To whom correspondence should be addressed

offspring.(24) TR679, which also has a high spontaneous mutation rate, was kindly supplied by Phil Anderson, Madison, WI and has been described previously.(25)

Isolation of pCW1

In a search for *unc-80* alleles, RW7097 was exposed to low concentrations (2% in air) of halothane (Halocarbon Lab, Inc., Hackensack, NJ) and screened for immobilized animals. Those immobilized individuals that regained mobility when removed from the anesthetic were outcrossed to N2 in order to fix the mutation. The new mutations were then mapped. Those not complementing *unc-80* were used for further studies.

We identified a spontaneous *unc-80* mutant (*fc2*) in the Bergerac hybrid strain, RW7097. *fc2* was outcrossed to N2 twelve times, with the *unc-80* phenotype reselected after each cross. During the seventh outcross, a spontaneous revertant was isolated (*fc52*). Genomic DNAs of N2, *fc2* and *fc52* were cut with BglII and probed with radioactively labeled Tc1. We identified a 3.2 kilobase (kb) fragment, which contains a Tc1 insert presumably responsible for the *Unc-80* phenotype. We cloned this fragment into pBS and named it pCW1.

cDNA library

A cDNA library isolated from N2, the wild type strain of *C. elegans*, was kindly provided by Bob Barstead, Washington University, St Louis. The library was made by adding EcoRI linkers to size-selected cDNAs (500bp–5000bp) and ligating into Lambda Zap (Stratagene).

Hybridizing methods

DNA was isolated as described by Maniatis *et al.* (26.) Genomic DNA was digested using restriction endonucleases (provided by Boehringer Mannheim, Inc.) and run on 0.8% agarose gels. Agarose was provided by Seakem. The DNA was transferred to Nytran filters by the method of Southern *et al.* (27) and fixed by exposure to ultraviolet light. (Stratalinker, Stratagene, Inc.)

Radioactive probe was made using ³²P labeled ATP in the random priming technique, (28) and run over a Sepharose G-50 column. The filters were prehybridized in Church's buffer at 68°C 4–6 hours and hybridized overnight in the same buffer at 68°C. Filters were then washed under stringent conditions as described by Maniatis *et al.* (26).

Sequencing methods

All clones were placed in pBS(+) and sequenced by the technique of Sanger and Coulson. (29) Initially, oligos to the T7 and T3 promoters were used, and oligos were synthesized to internal sequences for extension of the sequences. All *mariner* sequences were checked by sequencing both the (+) and (–) strands. Sequences were analyzed using the blast algorithm by NCBI. (30,31)

RESULTS

pCW1 contains a 300 base pair (bp) fragment on the 5' side of the Tc1 insert and a 1300 bp fragment on the 3' side. The 300 bp fragment and the 1300 bp fragment were subcloned into PBS and named pCW5 and pCW2 respectively. Both were used to probe genomic preparations from N2. pCW2 gave a unique pattern at 68°C but pCW5 appears to be repeated in the genome 15–20 times (Figure 1).

Figure 1. Southern blot of wild-type (N2) DNA digested with three different restriction enzymes [HindIII (a), EcoRV (b), and BglII(c)] and a spontaneous *unc-80(fc2)* DNA cut with BglII (d). This blot was probed with the 300 bp fragment, pCW5. This figure shows that pCW5 contains a sequence repeated 15–20 times.

The sequence of pCW5 showed homology to the *D.mauritiana* transposon, *mariner*. Since pCW5 consisted of only about a 300 bp fragment, while *mariner* consists of about 1300 bp, we wished to isolate the entire *mariner*-like (MLE) element. Using the unique fragment pCW2 as a probe, we isolated a 9.6kb fragment from N2. This larger fragment was cloned into the lambda-derived vector EMBL4, and called E4001. E4001 contained the same repetitive element as pCW5, as well as contiguous repetitive DNA. E4001 was subcloned and used to complete the sequencing of the MLE. This MLE was found to have 1017 base pairs with 50% homology to the *mariner* of *D.mauritiana* at the nucleic acid level and a 57% homology at the amino acid level (EMBL accession number X77804).

We screened a cDNA library from N2 for transcripts of the MLE. We found one full length cDNA which is nearly identical to the genomic MLE near *unc-80*, and contains flanking sequence on both sides (Data not shown). This clone, pCWZ2, contains 1330 bp, has 51% homology to the *mariner* of *D.mauritiana* at the nucleic acid level, and 58% homology at the amino acid level (EMBL accession number X77804). The beginning of translation in the cDNA appears to be at bp 258, and has a 21 bp deletion starting at position 924 compared with the *D.mauritiana* copy. The cDNA contains a 28 bp inverted repeat, but we have not found such a repeat in the genomic copy.

pCWZ2 differs by two nucleotides in the coding sequence compared to the genomic MLE, and by several nucleotides in the flanking sequences. The second change in the apparent coding sequence found in the genomic clone converts the stop codon to one coding for tryptophan. Although pCWZ2 most closely resembles the MLE of *D.mauritiana*, it represents a new subfamily of MLEs (Hugh Robertson, Personal Communication).

Figure 2. Comparison of the banding patterns of (a) *C. elegans*, (b) *C. briggsae*, (c) *C. remanei*, and (d) *C. vulgaris* when digested with the restriction enzyme EcoR1 and probed with the MLE pCWZ2. Note that the latter three show one band at about 6 kb at high stringency. No increase in banding was noted at low stringency.

Figure 3. Comparison of the banding patterns of N2 isolated seven years apart when digested with the restriction enzyme EcoR1 and probed with the MLE, pCWZ2 at high stringency. This length of time represents 500–600 generations of *C. elegans*. Lane 1 corresponds to a N2 culture from worms frozen in 1978 and unfrozen in 1992. Lane 2 corresponds to a N2 culture which has been unfrozen since 1985.

Figure 4. Comparison of the banding patterns of N2 with a) TR679 and b) Bergerac (B0) when digested with the restriction enzyme EcoR1 and probed with the MLE, pCWZ2 at high stringency. TR679 is a strain selected for a high spontaneous mutation rate associated with mobilization of several transposons. Bergerac (B0) is also noted for its high rate of mobilization of transposons, particularly Tc1. Note that few differences exist between these strains but that in three such cultures some changes can be noted in the banding patterns.

The MLE was used to probe a Southern blot comparing genomic DNA from *C. elegans* with the closely related species *C. briggsae*, *C. remanei*, and *C. vulgaris*. In each of the three related species, one band was noted at about 6kb when the DNA was digested with EcoR1, then probed with *mariner* and washed at high stringency (Figure 2). This band was noted in several other similar Southern blots, and was consistently decreased in intensity compared with bands in *C. elegans*. No other bands were

noted at lower stringencies, nor did the single band increase in brightness. (Data not shown.)

We obtained N2 worms frozen since 1978, and probed their DNA with the plasmid pCW5. We found that no change in banding pattern has occurred since that time when we compared the 1978 animals to those in culture, at present, in our lab (Figure 3). This length of time represents about 500 generations in *C.elegans*. It appears that, at least in N2, this element is not mobile.

We also compared the banding patterns of N2 with B0 and TR679 (Figure 4a, b). We grew several parallel cultures of B0 and TR679 to see if any of the MLEs mobilized. Genomic digests of each culture were followed by Southern blotting and probing with pCW5. We found that the overall banding pattern was very similar between B0, TR679 and N2. However, a small number of bands differed between independently grown cultures of B0 (Figure 4b), and these also differed from those of N2 (compare Figure 3 and 4b). We have not identified a new insertion of an MLE in *C.elegans*.

DISCUSSION

We have identified a repetitive DNA sequence in *C.elegans* found previously in insect species and planaria. (1,2) This MLE shows strong homology to the original *mariner* found in *D.mauritiana*, as well as MLEs found in other insect species. In addition, a separate *mariner*-like sequence has been identified during the *C.elegans* sequencing project. (GenBank accession number M98552) In *D.mauritiana*, *mariner* codes for a transposase that is capable of producing mutations by mobilizing other copies of *mariner*. (32) The cDNA of the *C.elegans* MLE is homologous to the *Drosophila* transposase, has an open reading frame coding for 339 amino acids, and has inverted terminal repeats. However, since *mariner* appears immobile in N2 for over 500 generations, it is unclear if the potential protein product of the cDNA still possesses any transposase activity. The genomic copy of *mariner* that we sequenced may be immobile due to loss of the inverted terminal repeats. In addition, the 21 bp deletion in the coding region of the nematode MLEs compared to those in *D.mauritiana* may render them wholly or partially inactive as transposases. This could lead to fixation of MLEs in the genome. A variety of single base changes can render the active *mariner* element inactive. (33,34) However, maintenance of an open reading frame coding for 339 amino acids tends to point to an active role for *mariner* in *C.elegans*. While it is clear that *mariner* does not mobilize in N2, Southern blots from B0 and TR679 leave open the possibility that *mariner* may move in these strains. Whether the RFLPs seen in these strains is due to intrinsic mobility of the MLE, or to other rearrangements, like Tc1 insertions, is not known. We have not identified such a mobile MLE in *C.elegans*, although others in the worm community are now attempting to do so. Thus, at the present time, *mariner* can not be called a transposable element in *C.elegans*.

We also note that the form of the cDNA from *C.elegans* is unusual. If this isolate represents a cDNA from the element itself then the promoter must lie outside the element. This seems unlikely and thus this cDNA probably arises from a coincidental insertion of the element near a promoter. Alternatively, activation of *mariner* in *C.elegans* may rely on such a placement to ensure its transcription and ultimate translation. No promoters are noted in either isolated copy of *mariner*. If they code for a functional transposase at least one copy must have a promoter. pCWZ2 may

represent such an MLE. If so, the characteristics of the promoter will govern when such a transposase is made, and may alter the time of transposition. It may be that *mariner* is mobile in tissues other than the germ line in *C.elegans*.

We originally isolated this element in a strain, RW7097, that causes another transposon, Tc1, to mobilize at a high rate. Our MLE was immediately adjacent to a Tc1 insertion causing a mutation in the *unc-80* gene. The *unc-80* gene is a relative 'hot spot' for transposon insertion, as we have many alleles of *unc-80* via spontaneous mutation of RW7097. It is not known if the presence of an MLE is associated with an increased frequency of Tc1 insertion. However, as noted by Henikoff and Henikoff (35), the putative transposase for *mariner* and Tc1 share a distant common ancestor. The structure of the local region may make it a 'hot spot' for both types of elements, or the presence of *mariner* in this region may make the DNA more accessible to a Tc1 insertion.

mariner is present in roughly equal copy number in geographically distinct strains of *C.elegans*, N2 and B0. In addition, the elements appear to be distributed in a similar manner on Southern blots in both strains. This may represent horizontal transfer of an MLE into a common ancestor of B0 and N2. An interesting, though remote, possibility is that the horizontal transfer could have occurred into both B0 and N2 very recently. The laboratory which isolated *mariner* from *D.mauritiana* and the laboratory which isolated the hybrid *C.elegans* strain, RW7097, and from which we had obtained our B0 strain, had been in close physical proximity for over ten years. We were interested to know if the MLE was present in *C.elegans* prior to the two groups arriving in St Louis. Since not all *mariner* elements have been sequenced in *D.mauritiana*, it seemed possible that the transfer may have occurred between the two laboratories. The similar banding patterns seen in N2 and B0 would then reflect preferred insertion sites. The presence of an identical pattern in N2 frozen in 1978 eliminates this possibility.

The presence of an identical 21 bp deletion in the genomic and cDNA copies of nematode MLEs compared with *D.mauritiana*, argues that these two copies descended from a common ancestor. The roughly identical banding patterns in B0 and N2 also indicates that these elements were fixed when B0 and N2 diverged. Taken together these facts indicate that B0 and N2 probably diverged after spread of MLEs throughout the genome. However, an MLE appears to be present in only a single copy in the closely related species, *C.briggsae*, *C.remanei*, and *C.vulgaris*. The similar sizes of the fragment containing the MLE in these species may represent a preferred insertion site of a recent horizontal transfer. A second possibility is that these species received multiple copies of the MLE from a common ancestor, but *C.briggsae*, *C.remanei*, and *C.vulgaris* underwent stochastic loss after *C.elegans* diverged. If so, one must explain why one element has remained. (Other copies may be present, but have diverged sufficiently to not be identified by low stringency hybridization.) Thirdly, it may be that the genus originally contained a single MLE which duplicated in the ancestral *C.elegans*. This duplication has been retained in both the N2 and B0 strains of this species, and has undergone faithful vertical transmission since the time of their divergence from one another (as indicated by the probing at high stringency). Since two subfamilies of MLEs have been identified in *C.elegans* (Genbank accession number M98552) (36), a separate horizontal transfer may also have occurred in this species. A fourth possibility is that the single MLE in the three species other than *C.elegans*

represents a truncated MLE with a sequence tightly conserved across the genus. This would explain the equally weak signal from our probe at both high and low stringency.

The presence of an MLE in a non-insect species raises questions as to its mode of dispersion between distant species, and its true extent among different phyla. The presence of multiple elements in one species of *Caenorhabditis* while closely related species may only have one argues for recent horizontal transmission. However, it is not impossible that *mariner* is truly ancient, and that tight conservation of sequences within diverse organisms has been necessary to maintain its function. This would allow for the *mariner* elements to have been present before the separation of planaria, nematodes and later, insects. A phylogenetic survey of the extent of MLEs and the relationship of the *C.elegans* subfamilies to those of other species would give information as to their evolution and mode of transmission.

ACKNOWLEDGEMENTS

The authors wish to thank Hugh Robertson for invaluable discussions, sharing of data, and advice. John Collins, Dan Hartl and Helen Salz provided helpful discussions and suggestions. Susan Thomas provided excellent technical assistance. This work was supported by the National Institutes of Health awards # GM 41385 and # GM 45402, and Academic Anesthesiology, Inc.

REFERENCES

- Robertson, H.M. (1993) *Nature*, **362**, 241–5.
- Garcia-Fernandez, F., Marfany, G., Baquana, J. and Salo, E. (1993) *Nature*, **364**, 109–110.
- Voytas, D., Konieczny, A., Cummings, M. and Ausubel, F. (1990) *Genetics*, **126**, 713–721.
- Wilke, C.M., Maimier, E. and Adams, J. (1992) *Genetica*, **86**, 155–173.
- Daniels, S., Peterson, K., Strusbaugh, L., Kiddwell, M. and Chovnick, A. (1990) *Genetics*, **124**, 339–355.
- Wichman, H.A., VanDen Busche, A., Hamilton, M.J. and Baker, R.J. (1992) *Genetica*, **86**, 287–293.
- Henikoff, S. (1992) *The New Biologist*, **4(4)**, 382–388.
- Liao, L., Rosenzweig, B. and Hirsh, D. (1983) *Proc. Nat. Acad. Sci. (USA)*, **80**, 3585–3589.
- Levitt, A. and Emmons, S.W. (1989) *Proc. Nat. Acad. Sci. (USA)*, **86**, 3232–3236.
- Collins, J., Forbes, E., Anderson, P. (1989) *Genetics*, **121**, 47–55.
- Yuan, J., Finney, M., Tsung, N., Horvitz, H.R. (1991) *Proc. Nat. Acad. Sci. (USA)*, **88**, 3334–3338.
- Dreyfus, D.H. and Emmons, S.W. (1991) *Nucleic Acids Res.*, **19(8)**, 1871–7.
- Moerman, D. and Waterston, R. (1989) In Berg, D., Howe, M. (ed.), *Mobile DNA*. Am. Society of Microbiology, D.C., pp. 537–556.
- Abad, P., Quiles, C., Tares, S. and Piotte, C. (1991) *J. Mol. Evol.*, **33(3)**, 251–258.
- Rosenzweig, B., Liao, L. and Hirsh, D. (1983) *Nucleic Acids Res.*, **11**, 4201–4209.
- Harris, L., Prasad, S. and Rose, A. (1990) *J. Mol. Evol.*, **30**, 359–369.
- Harris, L., Baillie, D. and Rose, A. (1988) *Nucleic Acids Res.*, **16**, 5991–5998.
- Henikoff, S. and Plasterk, R. (1988) *Nucleic Acids Res.*, **16**, 6234.
- Jacobson, J.W., Medhora, M.M. and Hartl, D.L. (1986) *Proc. Nat. Acad. Sci. (USA)*, **83**, 8684–8688.
- Maruyama, K. and Hartl, D.L. (1991) *Genetics*, **128**, 319–329.
- Lidholm, D.A., Gudmundsson, G.H. and Boman, A.G. (1991) *J. Biol. Chem.*, **266**, 1518–21.
- Brenner, S. (1973) *Genetics*, **77**, 71–94.
- Eide, D. and Anderson, P. (1985) *Proc. Nat. Acad. Sci. (USA)*, **82**, 1756–1760.
- Mori, I., Moerman, D.G. and Waterston, R.W. (1988) *Genetics*, **120**, 397–407.
- Collins, J., Saari, B. and Anderson, P. (1987) *Nature*, **328**, 726–728.
- Maniatis, T., Fritsch, E.F., Sambrook J. (1986) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Southern, E.M. (1975) *J. Mol. Biol.*, **98**, 503–517.
- Fernberg, A.P. and Vogelstein, B. (1984) *Anal. Biochem.*, **137**, 266–270.
- Sanger, F., Coulson, A.R., Barrell, B.G., Smith A.I.H. and Roe, B.A., (1980) *J. Mol. Biol.*, **143**, 161–178.
- Gish, W. and States, D.I. (1993) *Nature Genetics*, **13**, 266–272.
- Altschul, S.F., Gish, W., Miller, W., Meyers, E.W. and Lipman, D.I. (1990) *J. Mol. Biol.*, **215**, 403–310.
- Garza, D., Medhora, M., Koga, A. and Hartl, D.L. (1991) *Genetics*, **128**, 303–310.
- Maruyama, K., Schoor, K. and Hartl, D.L. (1991) *Genetics*, **128**, 777–784.
- Medhora, M., Maruyama, K. and Hartl, D.L. (1991) *Genetics*, **128**, 311–318.
- Henikoff, S. and Henikoff, J.G. (1992) *Proc. Nat. Acad. Sci. (USA)*, **89**, 10915–10919.
- Robertson, H.M., Lampe, D.J. and MacLeod, E.G. (1993) *Nucleic Acids Res.*, **20**, 6409.