ACADEMIC PRESS

Molecular Phylogenetics and Evolution 25 (2002) 210–217

MOLECULAR PHYLOGENETICS AND EVOLUTION

www.academicpress.com

Short communication

Characterization and molecular phylogenetic analysis of *mariner* elements from wild and domesticated species of silkmoths

M. Dharma Prasad,[a] Dmitry L. Nurminsky,[b] and Javaregowda Nagaraju[a,*]

[a] *Laboratory of Molecular Genetics, Center for DNA Fingerprinting and Diagnostics, ECIL Road, Nacharam, Hyderabad 500076, India*
[b] *Department of Anatomy and Cell Biology, Tufts University School of Medicine, 136 Harrison Ave., Boston, MA 02111, USA*

Received 2 August 2001; received in revised form 2 April 2002

Transposable elements are powerful tools for studying molecular genetics as they serve as agents for chromosomal insertions, deletions, or rearrangements and are found to be maintained in a variety of the genomes. The *mariner* like elements (MLEs), first isolated from *Drosophila mauritiana* (Haymer and Marsh, 1986; Jacobson et al., 1986), are now known to be present in a wide range of animal species (see review Hartl et al., 1997b), and plants (Feschotte and Wessler, 2002; Jarvik and Lark, 1998). MLEs are characterized by the presence of an ORF coding for a transposase of about 350 amino acids, short inverted terminal repeats at the ends, and a TA duplication at the insertion site (Lohe et al., 1996). The MLE transposase contains two highly conserved motifs WVPHEL and YSPDLAP separated by about 150 amino acid motifs, as well as a specific D,D(34)D signature motif (Doak et al., 1994; Robertson, 1993).

The MLEs have been classified into several distinct subfamilies according to sequence similarities; elements from different subfamilies are typically 40–56% identical at the nucleotide level (Robertson and MacLeod, 1993). Gene transfer between species, a phenomenon known as horizontal gene transfer, appears to have played an important role in the evolution of MLEs. Horizontal transmission of MLEs is inferred from the occurrence of very similar transposon sequences in distantly related species, and from presence of different subfamilies of *mariner* elements in any particular species (Robertson, 1993). The extremely broad host range of MLEs, indicative of the host independence of the transposition process, has attracted interests because of the potential use of MLEs for genetic manipulations with insect species, with special emphasis on insects of economic importance (Kidwell, 1993). Recent studies demonstrated the potential of mariner-based transformation vectors for introducing exogenous DNA into the wide range of hosts, including fruitfly, mouse, chicken, mosquito, zebrafish, and leishmania (for recent review see Plasterk et al., 1999).

MLEs also present important issues from an evolutionary point of view. The vast majority of MLEs are not functional, because they contain multiple inactivating mutations such as deletions, insertions, and nucleotide substitutions (Lohe et al., 1997; Maruyama et al., 1991; Robertson, 1993). The only *mariner* elements demonstrated to be autonomous are the *Mos*I from *D. mauritiana* (Medhora et al., 1991) and closely related elements from *Drosophila simulans* (Capy et al., 1992). This apparent predominance of inactive MLEs prompted the presumption that mutational inactivation is an important part of the MLE life cycle within the species, which follows the initial invasion by horizontal transmission. These processes, along with the stochastic loss of inactive MLEs by random genetic drift, have been implicated as possible mechanisms underlying the curious distribution of MLEs among species (Lohe et al., 1995; Hartl et al., 1997a).

To explore the evolutionary biology and dynamics of the MLEs, we undertook a study of MLEs in the genomes of diverse silkmoths collected from various parts of the world.

The species used for mariner analysis included representatives of the domesticated silkmoth *Bombyx mori* (chromosome no. 28, mariner nomenclature, *Bmmar*), (Indian polyvoltine strain, Nistari) and its wild progenitor, *Bombyx mandarina* (27, *Bmamar*) and wild silkmoths of the *Saturniidae* family [*Antheraea mylitta*

* Corresponding author. Fax: +91-40-155479.
E-mail addresses: jnagaraju@www.cdfd.org.in, jnagaraju@lycos.com (J. Nagaraju).

PII: S1055-7903(02)00225-7

(31, *Anmmar*), *Antheraea assama* (15, *Anamar*), *Antheraea polyphemus* (30, *Anpomar*), *Antheraea yamamai* (31, *Anymar*), *Antheraea proylei* (49, *Anprmar*), *Antheraea pernyi* (49, *Anpmar*), *Philosamia cynthia ricini* (13, *Pcrmar*), and *Antheraea roylei* (30, 31, 32, *Anrmar*)], and *B. mori* parasite, *Exorista bombycis* (Indian Uzi fly) (8, *Ebmar*). Genomic DNA was isolated from the liquid nitrogen-frozen whole individuals crushed to fine powder as described in Nagaraja and Nagaraju (1995).

PCR was carried out as per Robertson (1993), the products amplified were cloned and on an average, three clones from each species were sequenced. The sequences were deposited to the GenBank under the Accession Numbers AF125219–AF125245 and AF212131–AF212136. We designated the silkmoth mariner elements as per the nomenclature suggested by Robertson and Asplund (1996). To avoid confusion, the MLEs of *B. mori* sequenced in the present study are named as *Bmmar3.1-3.3*, and the other two MLEs isolated earlier are *Bmmar1* (Robertson and Asplund, 1996) and *Bmmar2* (renamed from BmMLE) (Tomita et al., 1997). To see whether any of the *B. mori* MLEs are transcribed, we submitted *B. mori* MLE sequences for homology search against the EST database of *B. mori* available as the silkbase at http://www.ab.a.u-tokyo.ac.jp.

The conceptual translation products of different partial *mariner* elements were aligned with judicious introduction of frameshifts (except for *Bmmar3.1* and *Ebmar1.1*) to conserve the reading frame using Clustal W program (Thompson et al., 1994) and then manually

```
                .......100.......110.......120.......130.......140.......150.......160.......17
Bmamar1         GIIRYELLPPGSPIDSELYYEQLMRLKQEVEIKRPELINRRDV-VFHHDNARPHT#ILAIQQKL---RELGWEVLMHPP
Bmamar2         GIIHYGLLPPGRTIDSQLYCEQLMRLKQEVGRKRPEFINR#G--VFHHDNASPHT-SLASQQKL---RELGWEVLMHPP
Bmamar3         GVVHYSFLKSGQTITANIYYQKLHTMKEELAAKQPRLVSRFRP-LLLHANARPHT-AQQTTTKL---DELQVECL*HPP
Bmamar4         GIVH*KLLPPGRTIDSELYCEQLMRLKQEVGRKRLKLINRKGV-VFDHDNARPHI-SLATQQKL---RELGWEVLMHPP
Anamar1         I*WPNVA#I-----------QQL*TMMEKLAVKQLRLINLSSS-LLF#---RTTT-SQQTVTKL---EELLLDYLRHLI
Anamar2         I*WPNVA#I-----------QQL*TMMEKLAVKQLRLINLSSS-LLF#---RTTT-SQQTVTKL---EELLLDYLRHLI
Anamar3         GVLHYSLLRSSQTI#ADLYCE*FQ#RMEHLAVKLQRLVNLSNP-QLRQNNAKLHT-SRQTVAKF---EELRLECL*HPP
Anamar4         GVIHYSFLRSGQMISADLYNQQLQTMTEKLGIKQPRVVNLSSR-LLLQDNAKPHA-SQKSIAKL---EELWLQRFRYPP
Anamar5         GVILYSFPRSSQMLTADLYYQ*LQ#MLEKLFVTQPSLVNLFSP-LLLKNIARLIL-HDKWSLN*---RECG--WNRHPQ
Anmmar1         GVIH-IFLLNGVSITADV#C*KLNTMMEKLTHLQPALVNRSSP-#ALHNNARTHT-AQQTYPKL---QELELEVLRHPS
Anmmar2         GVIHHSFLPNGLSITADVYCEELNTMMEKLAHLQPALVNRSSL-LLLHDNAGPHT-AQPTVSKL---QDLGLKVLRHPP
Anmmar3         SVIHHSFLPNRVSITADVYCEDLNIMMEKLAHLQPVLVNRCS#----HDNALRHT-AQQTVSKL---QELGLEVLLHAP
Anpmar1         VVIHHIRLPNGVSITADVYCEELNTMMEKLAHHQPALVNRSSL-LLLHDNARPH#-AQQTVSKL---QELGLEVLCHP-
Anpmar2         VVIHHIFLPNGVSITADVYCEELNTMMEKLAHHQPALVNRSSL-LLLHDNARPH#-AQQTVSKL---QELGLEVLCAP-
Anpomar1        GPELVSCTSAS*DLARRFFVNNCKKNDKNANF#QPRLTNISSP-QLLQDNARSHT-SRKTV#KL---EELQLEYLRNPP
Anpomar2        GPELVSCTSAS*DLARRFFVNNCKKNDKMQTFKQPRLTNISSP-QLLQDNARSHT-SRKTV#KL---EELQLEYLRNPP
Anpomar3        GPELVSCTSAS*DLARRFFVNNCKKNDKNANF#QPGLTNISSP-QLLQDNARSHT-SRKTV#KL---EELQLEYLRNPP
Anprmar1        GVIHHSFLPNGVSITADVYCEELNTMMEKLAHLQSSLVNRSSP-L#LHDNARPHT-AQQTVSKL---QELGLEVLRHPP
Anprmar2        GVIHHSFLPYGM#--ADVYCEELNTMMEKLAHLQPALANRLSP-LLLHDNARPHS-AQQTVSKL---QELGLEVLRHPS
Pcrmar1         GVILYSFPRSSQMLTADLY-YQLQTMLEKLFVKQPSLVNLFSP--LLKNIARPHT-SRQMVAKL---EGMRLE#-RHPQ
Pcrmar2         IMSFYRRVEPLIRIS---AILL**DQSKK#ERKRPE*TNRKGE-VFLHDNA-GHTHM*PFSIKR---AWLEGVCLSGR-
Pcrmar5         GVIYYEHLKSSETITDERYRTQLMRLSRALKEKRLQYNERHDKVILWHDNARPHV-AKVVKKYL---ETLKWEILPHPP
Anrmar1         GVIHHSFLPDGLSITADVYCEELNIMMEKLAHLQPALVNRSSS-LLLH#---PHT-AHQTVSKL---QELGLEVLRHPL
Anrmar2         GVIHHSFLPNGVSITADVYCEELNTMMEKLAHLQPALVNRSSP-LLLHDNARPHT-AQQTVSKL---Q*LRLEVLCHPL
Anrmar3         GVIHHSFLPSGESITADVYCEELNTMMEKLAHL*PTLVNRLSP-LLLHDNARPHT-VQQTVFKL---QELGLKVLRHPP
Anymar1         GVIHHSFLPNGVSITADVYCEELNTMMEKLALLQPALVNRSSP-LLLHANARPHT-AQQTASKL---QELGLEVLRHP-
Anymar2         GVIHHSFLPNGESITVDVYCEELNTMMEKLAHLQPALVNRSSP-LLLHDNARPHT-AQQTVSKL---QELGLEVLRPH-
Anymar3         GVIHHSFLPNGVSSTADVYCAELNTMMEKLAHLQPELVNR#SL-LLLHDNARPHT-AQQTVSKL---QELGLEVLRHP-
Bmmar3.1        GIIHYELLPISRTIDSELYCEQLMRLKHEVERKRPESINIRGM-VFHHDNARPHT-SLATQQKL---RELGWEVLMHPP
Bmmar3.2        GIIHYELLPPGRTIDSELYCEQLMSFKQELERKRPELINRRGV-VFHYDNARSHT-SLATQ*KL---R#LG*EVLMHP-
Bmmar3.3        GIIHYEL#PPGRIIDSELCCEQLMRLKQEVERKRTELINRRGV-VFHHDNARPHT-SLATQQKL---KELD*EVLMHPP
Bmmar2          GVVHYSFLKSGQTITADIYCQQLQTMKEELAAKQPRLVNYSRP-LLLHDNGRPHT-AQQTITKL---DELQLACLRHPP
Bmmar1          GVTEPYFCEKGIKTSAQVYQDTILEKVVKPLNN-TMFNNQEWS--FQQDSAPGHK-ARSTQSWLETNVSDFIRAEDWPS
Ebmar1          GIIFIDYLQKGQTINSDYYIALLERLKAEIVKKRPHLKKKKV--LFHQDNAPCHK-SGKTMAKI---YELGYELLPHPL
Ebmar2          GVVYFELLPRNRTINSDLYCEQLEKLGAAINEKRPELVNRKGV-IFHHDNARPHT-SLATRQKL---RELGWELLMHHH
D.mauritianaMos1 GVIYYELLKPGETVNRARYQQQLINLNRALQRKRPEYQKRQHRVIFLHDNAPSHT-ARAVRDTL---ETLNWEVLPHAA
M.destructormar1 GVLYYELLEPGQTITGDLYRTQLIRLKQALAEKRPEYAKRHGAVIFHHDNARPHV-ALPVKNYL---ENSGWEVLPHPP
D.erecta        GIIHYELLPYGQTLNSTIYCR*LDRLKQAIDQKRPELANRKGV-VFHQDNARPHT-SLMTRQKL---RKLGWEVLSHPP
H.cecropiaMLE   GVIHYSFLKCGQTITVDIYYQQLQAMKEELAAKHPRLVNRSRS-LLLHDNARPHT-AKQTTTKL---NKLQLECLRHPP
D.tigrinaMLE    GVIHYDFMVPGTSITSDVYCSQLDDMMEKLAIKQPKMFNRLTP-ILLHDNARPHS-AKNTVAKL---QQLGLETLRHPP
Hornfly3.1      GIIFIDYLEKGK#INSDYYMALLERLKVEIAAKRPHMKKKEV--LFHQDNAPCHK-SLRTMAKI---RELGFELLPHPP
A.gambiae22.5   GIFFIEYLQKHKIINSDYYKALLERLKVKSAAKRPHMKKKKV--LFHQDNAPCHK-SLRTMAKI---DELGFELLPHPP
A.atlasMLE      GVVHYSFLKSGLTITADVYCQQLQVMMEKLAAKQPRLINRSRP-LLLHDNARPHT-AQQTATKL---GELQLECLRHPP
A.melliferaMLE  EIVYFELLPPNRTINSVVYIEQLTKLNNAVEEKRSELTNRKVV-VFHHDDARPQT--SLVIRKL---LEFGWDVLPHSL
Insect cons.    G      L  G  I    Y   L         K P           HDNARPH     T  KL     L  E L HPP
Mariner family  G I  ELL  G TI    YC QL         KRP          FLHDNARPH     T  KL     LG E L HPP
```

Fig. 1. Multiple alignment of the conceptual translations of 490 bp fragments obtained by MAR124F and MAR276R primers from different silkmoths and from Uzi fly. Some of the published MLE sequences are also included as representative of different MLE subfamilies. (–) indicates gaps introduced to maintain alignment, (*) indicates stop codons, and (#) indicates frameshifts applied to maintain a reading frame that aligns with other MLEs. The insect consensus is derived by majority rule from all the sequences indicated in the figure. The *mariner* family consensus is from Robertson and Asplund (1996).

```
                1.......10........20........30........40........50........60........70........80........90
Bmamar1         --TERNLMNR#LIYDS-LLRRNETVPFLKKLITGDEK*ITY-------DKNVRKRSGSKAGQASQTVAKTGLIRN-KVMLCV#VW*-D*K
Bmamar2         --TERNLLNSVLNCDSLL#RRNETEPFLKKLITGDEKWITYNISLITYDKNVRKRSRSKAGQASQTVT*--L#RN--KAMLCVWW--DWK
Bmamar3         --SESNLQIRVECC-VTLLNRHNNEGILNRIITCDEK*ILY---------YNRKRSSQWLNPGAPAKSCPKLKFTQKKILVSVWW--TSA
Bmamar4         --IETNLINRVLICDSIL-RRNEAKPFLKKLITDDEKRITY-------DKYVRKRS--WSKAGQASQTVGKPGLTRNKVMLCV*W--DWK
Anamar1         --TEAYQQTHIKYY-VILLDHHNNEGFLDCILTCDKKCIFY---------DNRRRSSQWLNPGEPTKPCSK*KLTQKKVLMC--------
Anamar2         --TEAYQQTHIKYY-VILLDHHNNEGFLDCILTCDKKCIFY---------DNRRRSSQWLNPGEPTKPCSK*KLTQKKVLMC--------
Anamar3         --TETHQQTGIESS-ITLFNRYNNDGIL---------KWIFY---------NKRQHSSQWLNPGDLAKSCHKRKLTQKNILVS-VW--TCD
Anamar4         ---VKAYQTHIECR-VTLRNCQNNKEIFNRIVF#DEKWILY---------DNRKRSSQRLNPG*QIKCCPKL*PSSKKVLMSVRE--TST
Anamar5         --TEANQQTRIECI-VTLLNRYNYRNFT---------EWILY---------DNC--IALTHIIGEPA#-PVQRK*TQKKILVS#WW--TCA
Anmmar1         --NDRQREVRVKTC-LALLNRHTNEEILNRIVTCDEKWILF---------DNRKCSASWLDSGSALKQCPKRKLTSRKVMVTVWW--SSA
Anmmar2         --NNRQREVRVETC-PGLLNRHTNEGILNCIVTCDKKWILF---------DNRKRSASWLDPGSAPK*CPKRKLTPRKVMFTGWW--SSA
Anmmar3         --Q*SQREVRVETC-LVLLNRHTNEGILNLIVTCDEKLILF---------DNRKRSASWLDPGSAPKQCPKRKLTPRKVMVTVLW--SSA
Anpmar1         --NDRQREVRFETC-LALLNRHTNEG*INRIVTCYEKWILF---------DKRKRSAS#LHPGSAPKQCPIRKLTPRKVMFTVWW--ASA
Anpmar2         --NDRQREVRFETC-LALLNRHTNEG*INRIVTCYEKWILF---------DKRKRSAS#LHPGSAPKQCPIRKLTPRKVMFTVWW--ASA
Anpomar1        --TKAHQPTRIECC-VVLINRHNNNGILNHIVTCDVK*ILY---------DNWKLT-SQNRRNSVPSEC-----WLRKRY*GVF-----V
Anpomar2        --TKAHQPTRIECC-VVLINRHNNNGILNHIVTCDVK*ILY---------DNWKLT-SQNRRNSVPSEC-----WLRKRY*GVF-----V
Anpomar3        --TKAHQPTRIECC-VVLINRHNNNGILNHIVTCDVK*ILY---------DNWKLT-SQNRRNSVPSEC-----WLRKRY*GVF-----V
Anprmar1        --NDRQLEVRVETC-LALLNRHTNEGILNRIVTCDEKWILF---------DNRKRSASWLDSGSAPKQCPKRKLTPRKMMVTVWW--SSA
Anprmar2        --LSAQREVRVETC-LA#LNRHTNEGIFNSIVTCDEKWILF---------DNRKCSASWLDPGSAPKQCPKRKLTPRKMMVT#WW--SSA
Pcrmar1         --TEANQQTRIECF-VTLLNRYN*QG#F*L------KWILY---------DNCKRSS#THIIGEPAKSGPRRKLTQKKILVSVWW--TCA
Pcrmar2         --TESNLINRVLK#----------EPFLKRLITDDMKWTRT---------TN#KRSKVG----QASQTVAKRGLTRDKVMLCVWW-N*LF
Pcrmar5         --NRGDIE*RLFAC-EQLLERQTRKGFLHRIVTRDEKWVHY---------NNPKRRKSWRLPGHTATSTPRPNIHASKVMLCV*W--AQL
Anrmar1         --NDRQHEVRVETC-LALINRHTNEGILNRIVTCDEKWILF---------DNRKRSASWLV#GSAPKQCSKRILIPRKVMVTGWW--SSA
Anrmar2         --NDRQREVRVETC-LALLNRHINEGILSRIVMCNDKWILF---------DNRKRSACWLDPGSAPKQCPKRKLTPRKVMVTVWW--SSA
Anrmar3         --NDRQCEIRVETC-FALLDRHTNEEILNRIATCDVKWFLF---------KNRKRSVSWLDPSSAPKQCPKRKLIPRKVMVTVWW--SSA
Anymar1         --NDRQREVRVETC-LALLNRHTNEGILNRIVTCVERWILF---------DNRKRSASWLDPGSALKQCPKRKLTPRKVMVTVWW--FRA
Anymar2         --NDRQREVRVETC-LALLNRHTNEGILNRIVTCDEKWILF---------DNRKRSASWLNPGSASKQCPK*KLTPR#---TVWW--SSA
Anymar3         --NDRQREVRVETC-LALLNRHTNEGILNRIVTCDEKWILF---------DNRKRSASWLDPGSAPKQC#KRKLTPRKVMVTVWW--SSA
Bmmar3.1        --TERNIMNRVLICDS-LLRRNETEPFVKKLITGDEKWITY-------DKNVRKR--SWSKAGQASQTVTKPGLTRNKAMLCVWW--DWK
Bmmar3.2        --TERNLMNRVLICDS-LLRRNETEPLLKKLITGDGKWITY-------DKNVRKR--SWAKVGQASQTVTKPGLTRNKVMLCVWW--NWT
Bmmar3.3        --TEVTE#-RVLICDS-LLRR*NTEPF*KKLITGDEKWIMY-------DKNVRKR--SWSKAGQASQTVAKPGLTRNKVVLYVWW--DWK
Bmmar2          --SESNLQTRIDCC-VTLLNRHYNEGILNRIITCDEKWILY---------DNRKRLSQWLNPGDPAKSYPKRKLTQKKLLMSVWW--LSA
Bmmar1          --TDNLKENRVVKS-KQLLKRYAKGG-HRKILFTDENFFTI----EQHFNKQNDRIYAQSSKE-ASQLVDRVQRGHYPTSVMVWWGISYE
Ebmar1          TIDQKQ*RVDDSEQCLALFKGNKAEFFR*-YVTMDETWLHHFTP-----ESNRQSS-EWTASD--NPKIGKTQQSAGKVMASVFW--DAH
Ebmar2          --KEIHLTNRFNACDIHL-KRYEDDPFLKQIITGNEKWTVY-------NNVNRKRS--WSKHDEAPQTTSKADIQQKKVMLSVWW--DWK
D.mauritianaMos1 --NERQMERRKNTC-EILLSRYKRKSFLHRIVTGDEKWIFF-------VNPKRKKS--YVDPGQPATSTARPNRFGKKTMLCVWW--KQS
M.destructormar1 ---KPRDVERRFCMSEMLLQRHKKKSFLSRIITGDEKWIHY-------DNSKRKKS--YVKRGGRAKSTPKSNLHGAKVMLCIWW--DQR
D.erecta        --TQKNLLDRINACDMLL-KRNELDPCLKRMVTGDEKWITY-------DNIKRKRKR------ESSQTVAKPGLTARKVLLCVWW--DWK
H.cecropiaMLE   --SESNLQTRVD-CYVTLLNRHNNE----------R-KMY---------DNRKGSLQWLNPGDPAKSCPKR*LTQKKLLVVFWW--TSA
D.tigrinaMLE    ---DEHKQQRLDAC-LSLLSRNKADPFLHRIVTCDEKWIMY---------DNRKRSSQWLDPDEPPKHCPKRKVHQKKLMVTVWW--SSY
Hornfly3.1      --QKQQRVDNSERC-LQLLTRNTPEFFRR-YMTMDETWLHHYTP-----ESNRQ-SAEWTATGEPSPKRGKTQKSAGKVMASVFW--DAH
A.gambiae22.5   --QKHQRVDDSERC-LQLLTRNIPE-FLRRYVTMGETWLHHYTP-----ESNRQ-SAQWTWTGEPAPKRGKTQKSAGKVMASVFW--DAH
A.atlasMLE      -----NRQTRVDCC-VTLLNRHNNEGILNRIITCDEKWILY---------DNRKRSSQWLNPGEPAKSCPKRKLTQKKLLVSVWW--TSA
A.melliferaMLE  --KEKHLMQRINSC-DLLKKRSRNDPFLKRLITGDEKWIVY-------NNIKRKR--WWSKPREPAQTTLKTGIHQKKVLLSVWW--DYK
Insects-cons.           R   C   LL R      L    T DEKWI           NRKRS  W          K      K    VWW
Mariner family       QK  R   C   LL R     FL R VTGDEKW  Y        N  RK  W    E       K   H KK M  VWW  D
```

Fig. 1. (*continued*)

edited to optimize the alignment (Fig. 1). The full-length sequence alignments (data not shown) were performed manually with the help of an eyeball sequence editor. In some of the elements, including *Anamar1.1* and *1.2* and *Anpomar1.1, 1.2*, and *1.3,* large-scale deletions at different positions had occurred, whereas *Bmmar1.2* had a significant insertion of 21 bp. The similarity between sequences ranged from 20% to 85%. A number of blocks of amino acids, such as TCDEKWI, NRKRS, VWW, HDNARPH, and HPP, are highly conserved in insects (Fig. 1) and apparently define the functional transposase regions. Among these, consensus residues DEKW and HDNARPH have been proposed to be linked with catalytic activity of the *mariner* transposase (Doak et al., 1994).

A phylogenetic tree rooted with *Bmmar1* as outgroup was generated by the PHYLIP program package using the parsimony method with 1000 bootstrap replications (Ver. 3.572c, Felsenstein, 1993). A similar tree was obtained based on the distance matrix to indicate the branch lengths. The tree obtained was compressed by ignoring the branches with confidence values below 50%. Some of the published sequences were used as representatives from different subfamilies (Robertson, 1993) and we classified elements into five subfamilies as described by Robertson and MacLeod (1993) (Fig. 2).

Genomic DNA from *A. mylitta* was used to construct a genomic library. The library was screened with a 450 bp *A. mylitta mariner* fragment generated by PCR as described earlier. Three positive plaques were randomly selected, and their inserts excised and sequenced. The sequences obtained are submitted to the GenBank with Accession Numbers AF126011, AF126012, and AF125249. The full-length sequence alignment was performed manually. None of the three sequenced clones had a functional ORF because of stop codons, frameshifts, and deletions. The signature sequence D,D(34)D, characteristic of MLEs, was identified in the sequences, although in two clones (*Anmmar4* and *Anmmar6*) the last aspartic acid residue was deleted (data not shown). The inverted terminal repeats were found to have conserved blocks of TAGGT and TATGAA, which may play a role in recognition, by transposase. The genomic sequences flanking three full-length

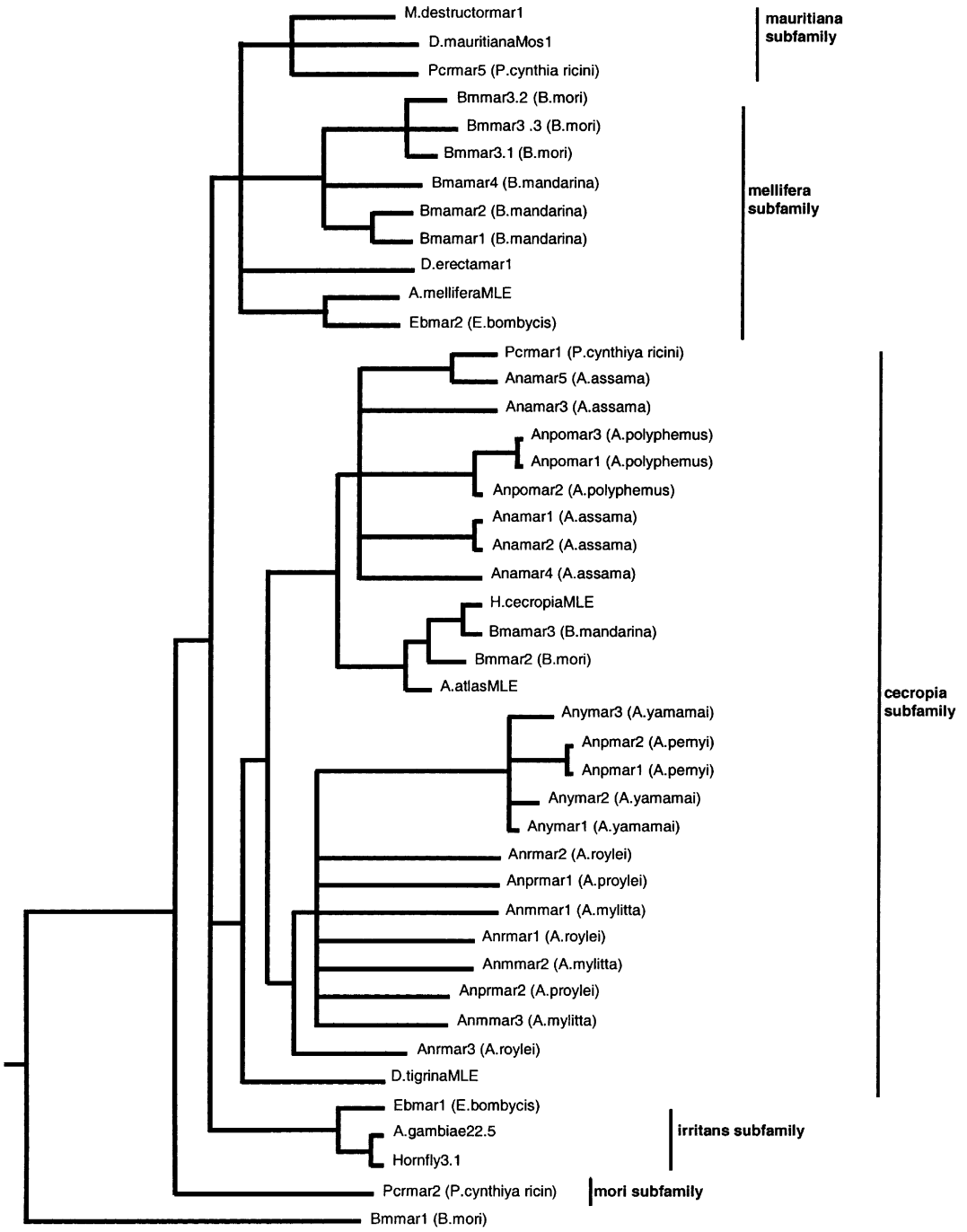

Fig. 2. Dendrogram showing the phylogenetic relationships between the conceptual translation products of partial *mariner* elements. The consensus tree is generated using protein distance and Fitch-Margoliash methods of PHYLIP program. Branches with lower than 50% confidence values after 1000 bootstrap replicates in Protein Parsimony method of PHYLIP were ignored. The elements are classified into five subfamilies based on their segregation into clusters.

A. mylitta MLEs do not show any consensus, although all of them are AT rich regions. A direct TA duplication at the site of insertion of the transposable element was also identified.

The copy numbers of the *mariner* elements in *B. mori*, *B. mandarina*, *A. mylitta*, *A. assama*, *A. proylei*, *A. roylei*, and *P. cynthia ricini* were estimated by dot blot analysis. Serial dilutions of genomic DNA and the cloned partial MLE segments from the same species (*Bmmar3.1*, *Bmamar1.1*, *Anmmar1.1*, *Anamar1.1*, *Anprmar1.1*, *Anrmar1.1*, and *Pcrmar1.1*) were spotted onto membrane and hybridized with their respective mariner elements at high stringency. The optical densities of signals from the cloned MLEs were used for estimating the copy numbers of MLEs in silkmoth genomes. The copy numbers were calculated assuming the haploid genome size of *B.*

mandarina and other wild silkmoths as equivalent to the *B. mori* genome size of 540 Mbp (Gage, 1974; Rasch, 1974). Under this assumption, the genomes of *A. proylei*, *A. mylitta*, and *A. roylei* contain 1500, 2000, and 5000 copies/haploid genome, respectively, whereas *A. assama* and *P. cynthia ricini* contain fewer MLEs of about 70 and 250, respectively. *B. mandarina*, an immediate ancestor of *B. mori*, harbors about 600 copies of *Bmamar1.1*. *B. mori* carries approximately 900 copies of *Bmmar3.1* which is lower than the 2400 copies of the basal element, *Bmmar1* reported earlier by Robertson and Asplund (1996) and higher than 90 copies of *Bmmar2* characterized by Tomita et al. (1997) from the *B. mori* genome.

The molecular evolution of synonymous and non-synonymous substitutions in the coding region was analyzed using the Codeml of PAML3.1 (Yang, 1997). Patterns of variation in the *Ka*/*Ks* ratio across the transposase gene were calculated by dividing the gene into 17 non-overlapping sections of 20 codons each. *Anmmar6* sequence was used as a reference and compared with homologous regions in DmMos1 (*D. mauritiana*), *Anmmar5*, HcMLE (*H. cecropia*), *Bmmar2*, and AaMLE (*A. atlas*) (Fig. 3). The *Ka*/*Ks* values (vertical axis) are plotted against the position of segment in the transposase ORF (horizontal axis). The two conserved regions corresponding to the active site of the transposase clearly show the lowered *Ka*/*Ks* values, as compared to the fluctuating higher values in the non-conserved regions. The low *Ka*/*Ks* values indicate that observed conservation of amino acid blocks is a result of purifying selection, apparently directed at the maintenance of functional domains of transposase. It is noteworthy that the same profile of *Ka*/*Ks* was observed after comparisons of the MLEs from different subfamilies (*Anmmar6* vs DmMos1, *HcMLE*, *BmMLE*, and *AaMLE*), and after comparison of the two MLEs that

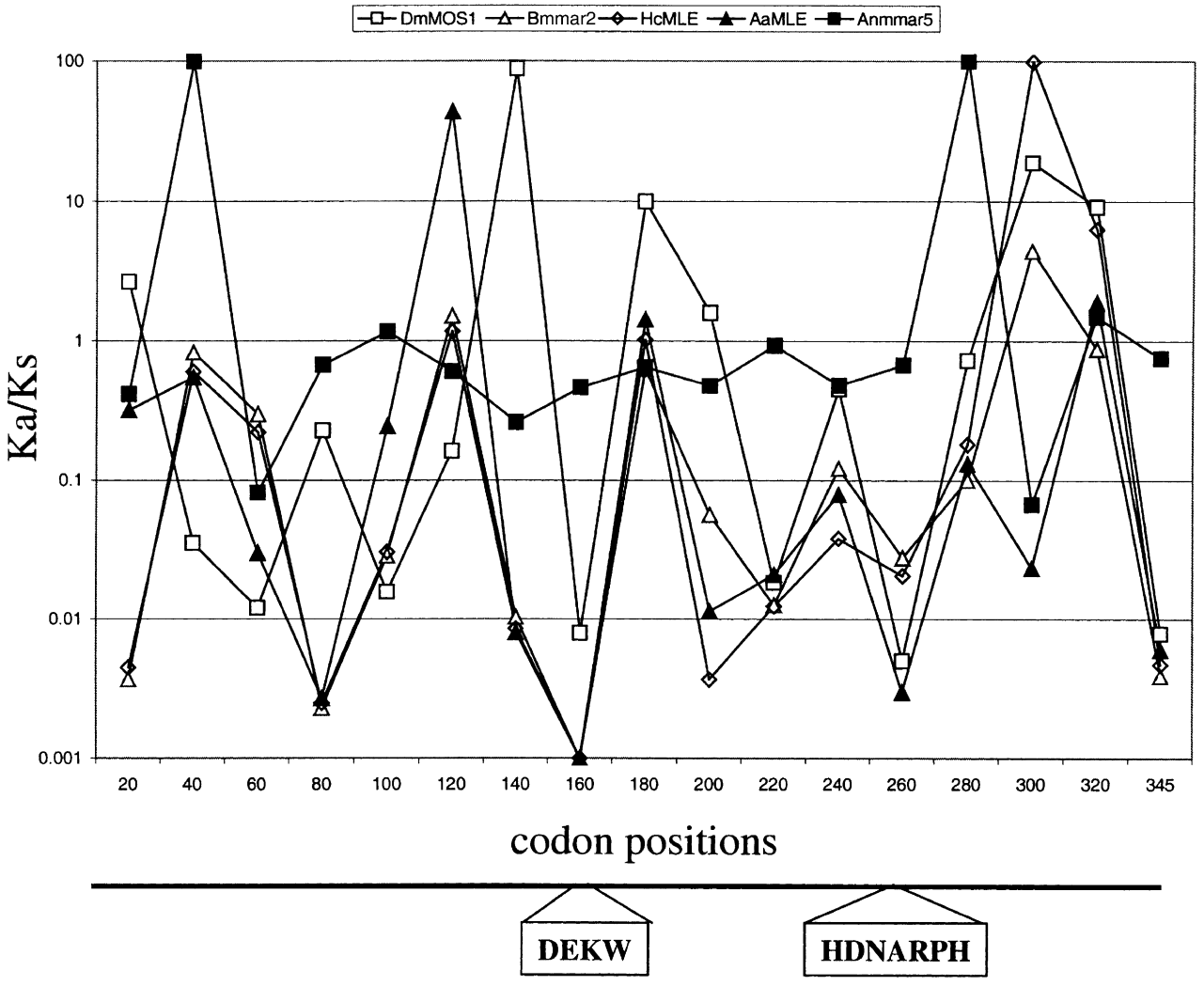

Fig. 3. Distribution of the *Ka*/*Ks* value across the transposase genes. The *Anmmar6* element was compared with *Anmmar5*, DmMos1, HcMLE, BmMLE, and AaMLE *mariner* elements. The values obtained are plotted in a logarithmic line graph. Vertical axis indicates the *Ka*/*Ks* values and horizontal axis the codon positions. Each data point represents 20 codons. The line below the graph represents transposase protein with conserved sequences shown in boxes.

belong to the same *cecropia* subfamily isolated from the same host *A. mylitta* (*Anmmar6* vs *Anmmar5*). This observation indicates that a reasonable part of the divergence of MLEs within *A. mylitta* lineage occurred under selective pressure for functional transposase, implying current or at least recent MLE transpositions in *A. mylitta*.

The phylogeny and pattern of divergence of MLEs in silkmoths were examined in relation to the phylogeny of the host species. Almost all the silkmoth MLEs contain conserved features that are characteristic to MLEs such as the D,D(34)D motif (Doak et al., 1994; Lohe et al., 1997). Out of 33 partial MLEs sequenced 31 were defective due to stop codons or frameshifts in the transposase ORF. So also was the case with the three copies of the full-length mariner elements isolated from *A. mylitta*. Our results indicated that, in general, phylogenetic relationships between MLEs obtained from diverse silkmoths are similar to the phylogeny of the host species consistent with the vertical inactivation stage of the MLE life cycle. It looks probable that most of these elements were present in the ancestral lineage prior to the divergence of these species and neutral evolution has occurred independently in each copy with respect to coding of amino acids in the transposase gene. For example, the MLEs from *Antheraea* species, *A. roylei*, *A. pernyi*, *A. proylei*, *A. mylitta*, and *A. yamamai* clearly belong to a subgroup of closely related elements within the *cecropia* subfamily (Fig. 2). Close relationship has been shown between *A. pernyi* and *A. yamamai* by molecular phylogenetic analysis (Shimada et al., 1995), and between *A. roylei*, *A. pernyi*, and *A. proylei* by interspecific hybridization (Nagaraju and Jolly, 1985).

Antheraea assama is considered different from the other *Antheraea* species, and more close to the common ancestor of *Antheraea* and *Philosamia*. Accordingly, the MLEs from *A. assama* comprise a separate subgroup within the *cecropia* subfamily, along with the MLEs from probably related species, *A. polyphemus*. These results imply that the MLEs from *cecropia* subfamily existed in the genome of the common ancestor of the *Antheraea* and *Philosamia* species.

Among *Bombycidae*, *B. mori*, and *B. mandarina* are very closely related species that diverged only about three million years ago (Maekawa et al., 1988), so are the MLEs obtained from these species. Majority of *B. mori* and *B. mandarina* MLEs belong to the *mellifera* subfamily. One of the *B. mandarina* MLEs detected in this study (*Bmamar1.3*) belongs to the *cecropia* subfamily, where it again clusters with *Bmmar2* (named earlier as BmMLE, Tomita et al., 1997). The existence of *cecropia* and *mellifera* types of MLEs in *mandarina* and *mori* genomes suggests that they predate the speciation event. Inherited by the diverged species, the MLEs continued to evolve as the parts of the species genomes.

According to hypothesis put forward by Hartl et al. (1997b), MLE is introduced into the host species by horizontal transmission, followed by a brief surge of active transpositions and subsequent down-regulation by overproduction inhibition. MLEs then enter the period of vertical inactivation, i.e., accumulation of mutations which render the elements defective. Further, consecutive or even simultaneous invasion of MLEs from different subfamilies results in the presence of divergent elements within the same species (Hartl et al., 1997b). Our analysis of silkmoth MLEs provides a number of examples of that kind. Both *B. mandarina* and *B. mori*, contain MLEs of two distinct subfamilies, *mellifera* and *cecropia*. In addition, *B. mori* contains an unusual element, *Bmmar1*, which is very different from all other MLEs. Analysis of MLEs from *Philosamia cynthia ricini* identified two different types of elements. One of these belongs to the *cecropia* subfamily, being closely related to the MLEs from *A. assama*. This could easily be expected, since the cytotaxonomical and interspecific hybridization studies have implicated that *A. assama* is close to the common ancestor of all other species of *Antheraea* and *Philosamia*. Another MLE from *P. cynthia ricini* is quite similar to the *Mos*I element from *D. mauritiana* and belongs to the *mauritiana* subfamily.

Horizontal transmission has been invoked to explain peculiar distribution of MLEs in the host species (Robertson, 1993). Although much speculated about, the mechanism of this process remains obscure. In the present study, we examined the MLEs from *Exorista bombycis* (Indian Uzi fly), a dipteran endoparasite of silkmoth *B. mori*, the sequence of which did not show any particular similarity to the silkmoth MLEs. Probably we may have to examine several mariner copies from this parasite to make inference on possible horizontal transmission.

We measured the copy number of MLEs in silkmoth species to evaluate the dynamics of the elements. *B. mori* genome has been found to harbor *mori*, *mellifera*, and *cecropia* MLEs in 2400, 900, and 100 copies, respectively. This diversity of copy number between different types of MLEs in *B. mori* may probably reflect different stages in the evolution of these elements after their invasion into the species at different time points, apparently undergoing mutational degeneration. However, we cannot exclude the possibility of presence of active copies in this species. One possible example is a *mellifera* subfamily MLE *Bmmar3.1*. Partial sequencing of this element revealed an intact transposase ORF and the silkworm EST database search identified a cDNA corresponding to this element. Although final conclusions must await the cloning and sequencing of the full-length *Bmmar3.1* and detailed transcription analysis, the data suggest that *Bmmar3.1* may represent a transcribed silkmoth MLE coding for functional transposase.

The abundance of the related MLEs differs dramatically in the different silkmoths examined. *Bmmar3.1* and

Bmamar1 are present in 900 and 600 copies in the genome of *B. mori*, and *B. mandarina* respectively. The genome of *A. assama* contains 70 copies of *Anamar1*, at the same time *P. cynthia ricini* harbors 250 copies of closely related *Pcrmar1*. Another observation is the difference in copy number of the *Anrmar1*/*Anprmar1* element between *A. roylei* (5000 copies) and *A. proylei* (1500 copies). *A. proylei* is a synthetic species derived from the interspecific hybrid between *A. roylei* and *A. pernyi* just 72 generations ago. The difference in the time of acquisition of the elements, continuing transpositions, and stochastic loss of the MLE copies by genetic drift could account for the copy number diversity in these species. For example, *A. assama* is confined to only small pockets of Assam state of India and has very narrow genetic variability (Nagaraju et al., Unpublished results). Under this assumption, the vertical transfer of MLE into *A. assama* appears to have been the most ancient and the genetic drift would have accelerated the depletion of copy number in *A. assama*. The drastic difference in copy number between *roylei* and *proylei* MLEs may not be surprising in light of the observations that selective elimination of the *A. roylei* chromosomes from the hybrid (*A. proylei*) occurs resulting in the loss of *A. roylei*-derived MLEs (Nagaraju and Jolly, 1985).

Analysis of the distribution of *Ka*/*Ks* ratio along the transposase ORF demonstrated much lower than average *Ka*/*Ks* values in the regions containing the conserved blocks of amino acids (Fig. 3). This observation indicates the purifying selection pressure directed at conservation of functional domains of transposase, implying that at least substantial part of the divergence between silkmoth MLEs occurred under selective constraint. The differences between MLEs isolated from different host species most likely have been acquired during multiple rounds of MLEs expansion that probably included horizontal transmissions followed by periods of active transpositions. As these events require the transposase activity, it is not surprising that the comparisons of MLE from *A. mylitta* (*Anmmar6*) with the MLEs isolated from different host species (*D. mauritiana*, *H. cecropia*, *B. mori*, and *A. atlas*) revealed the selective pressure for transposase conservation. What is more intriguing, comparison of the two MLEs, *Anmmar6* and *Ammar5*, that belong to the same subfamily and were isolated from the same host *A. mylitta*, demonstrated a similar pattern of conservation of transposase domains, indicating that a significant part of the MLE divergence within *A. mylitta* occurred under selective pressure as well.

Acknowledgments

The authors are grateful to Daniel Hartl, Harvard University for providing facilities to carry out some part of the work, and useful suggestions on the manuscript. The work was supported by grants from Department of Biotechnology, Government of India, New Delhi to J.N. M.D.P. is thankful for Council of Scientific and Industrial Research, for financial support.

References

Capy, P., Koga, A., David, J.R., Hartl, D.L., 1992. Sequence analysis of active *mariner* elements in natural populations of *Drosophila simulans*. Genetics 130, 499–506.

Doak, T.G., Doerder, F.P., Jahn, C.L., Herrick, G., 1994. A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "D35E" motif. Proc. Natl. Acad. Sci. USA 91, 942–946.

Felsenstein, J., 1993. PHYLIP 3.5, available at joe@genetics.washington.edu., University of Washington.

Feschotte, C., Wessler, S.R., 2002. Mariner-like transposons are widespread and diverse in flowering plants. Proc. Natl. Acad. Sci. USA 99, 280–285.

Gage, L.P., 1974. The *Bombyx mori* genome: analysis by DNA reassociation kinetics. Chromosoma 45, 27–42.

Hartl, D.L., Lozovskaya, E.R., Nurminsky, D.I., Lohe, A.R., 1997a. What restricts the activity of *mariner*-like transposable elements. Trends Gen. 13, 197–201.

Hartl, D.L., Lohe, A.R., Lozovskaya, E.R., 1997b. Modern thoughts on an ancyent *marinere*: function, evolution, regulation. Ann. Rev. Genet. 31, 337–358.

Haymer, D.S., Marsh, J.L., 1986. Germ line and somatic instability of a *white* mutation in *Drosophila mauritiana* due to a transposable genetic element. Dev. Genet. 6, 281–291.

Jacobson, J.W., Medhora, M.M., Hartl, D.L., 1986. Molecular structure of somatically unstable transposable element in *Drosophila*. Proc. Natl. Acad. Sci. USA 83, 8684–8688.

Jarvik, T., Lark, K.G., 1998. Characterization of *Soymarl*, a mariner element in soybean. Genetics 149, 1569–1574.

Kidwell, M.G., 1993. Lateral transfer in natural populations of eukaryotes. Ann. Rev. Genet. 27, 645–662.

Lohe, A.R., De Aguiar, D., Hartl, D.L., 1997. Mutations in the mariner transposase: the "D,D(35)E" consensus sequence is nonfunctional. Proc. Natl. Acad. Sci. USA 94, 1293–1297.

Lohe, A.R., Moriyama, E.N., Lidholm, D.-A., Haiti, D.L., 1995. Horizontal transmission, vertical inactivation, and stochastic loss of *mariner*-like, transposable elements. Mol. Biol. Evol. 12, 62–72.

Lohe, A.R., Sullivan, D.T., Hartl, D.L., 1996. Genetic evidence for subunit interactions in the transposase of the transposable element *mariner*. Genetics 144, 1087–1095.

Maekawa, H., Takada, N., Mikitani, K., et al., 1988. Nucleotide organizers in the wild silkworm *Bombyx mandarina* and the domesticated silkworm *B. mori*. Chromosoma 96, 263–269.

Maruyama, K., Schoor, K.D., Hartl, D.L., 1991. Identification of nucleotide substitutions necessary for trans-activation of *mariner* transposable elements in *Drosophila*: analysis of naturally occurring elements. Genetics 128, 777–784.

Medhora, M., Maruyama, K., Hartl, D.L., 1991. Molecular and functional analysis of the *mariner* mutator element *Mos*I in *Drosophila*. Genetics 128, 311–318.

Nagaraju, J., Jolly, M.S., 1985. Interspecific hybrids of *Antheraea roylei* and *A. pernyi*—a cytogenetic reassessment. Theor. Appl. Genet. 72, 269–273.

Nagaraja, G.M., Nagaraju, J., 1995. Genome fingerprinting in silkworm, *Bombyx mori* using random arbitrary primers. Electrophoresis 16, 1633–1638.

Plasterk, R.H., Izsvak, Z., Ivics, Z., 1999. Resident aliens: the Tcl/mariner superfamily of transposable elements. Trends Genet. 15, 326–332.

Rasch, E.M., 1974. The DNA content of sperm and hemocyte nuclei of the silkworm, *Bombyx mori* L. Chromosoma 45, 1–26.

Robertson, H.M., 1993. The *mariner* transposable element is widespread in insects. Nature 362, 241–245.

Robertson, H.M., Asplund, M.L., 1996. *Bmmar1*: a basal lineage of the mariner family of transposable elements in the silkworm moth, *Bombyx mori*. Insect Biochem. Mol. Biol. 26, 945–954.

Robertson, H.M., MacLeod, E.G., 1993. Five major subfamilies of *mariner* transposable elements in insects, including the Mediterranean fruit fly, and related arthropods. Insect Mol. Biol. 2, 125–139.

Shimada, T., Kurimoto, Y., Kobayashi, M., 1995. Phylogenetic relationship of silkmoths inferred from sequence data of the arylophorin gene. Mol. Phylogenet. Evol. 4, 223–234.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penaltied and weight matrix choice. Nucl. Acids. Res. 22, 4676–4680.

Tomita, S., Sohn, B.H., Tamura, T., 1997. Cloning and characterization of the mariner-like element in the silkworm, *Bombyx mori*. Genes Genet. Syst. 72, 219–228.

Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. CABIOS 13, 555–556, Available from. Available from <http://abacus.gene.ucl.ac.uk/software/paml. html>.