

The complete sequence of mag, a new retrotransposon in *Bombyx mori*

Jean-Jacques Michaille, S. Mathavan, Janine Gaillard and Annie Garel
Centre de Génétique Moléculaire et Cellulaire, CNRS UMR 106, Université Lyon I, 43 Boulevard du 11 Novembre 1918, F-69622 Villeurbanne cedex, France

Submitted December 21, 1989

EMBL accession no. X17219

Mag, a 4564bp long transposable element has been discovered in the large intron of a cloned allele of the Ser2 gene (1). A few copies (7 to 14) are dispersed into the genome of different strains of *Bombyx mori*. It is flanked by a 5bp repeat of the target sequence and is bordered by direct terminal repeats of 77 nucleotides. Two large open reading frames are organized as the gag and pol genes of retroviruses in the non coding strand of the Ser2 gene. The ORF1 is 258 codons long and presents the characteristic features of two nucleic acid binding motifs (underlined a). The ORF2 (1195 codons) shows strong homologies with the retroviral protease (b), reverse transcriptase (c), Rnase H (d) and endonuclease (e, f), in this order. The

unusually short terminal repeat, different from the arrangement of the LTR of retroviruses, has been confirmed by sequencing other copies of this element selected from a *Bombyx* genomic library. From the phylogenetic tree established on the RT sequence (2 and personal communication), this retrotransposon can be positioned among the copia like family elements of *Drosophila*.

REFERENCES

- 1. Michaille, J.J., Garel, A. and Prudhomme, J.C. (1990) *Gene* in press.
- 2. Xiong, X. and Eickbush, Th.H. (1988) *Mol. Biol. Evol.* 5, 675-690.

<pre> TGTATGTTG GTATGATGT ATTCAGCTCA GCTGTGATAC AACGTCGMA CGCATCTATTA TTTACTATGT GGGAGCGCG AAAAAAAAA TCATTTTGT GCTTTTGT TTTCTTTT GACTATTAH AHTAETATG CTGTGACAG TTAGAAATF TCGCGTBAH CACTGTAAH F P L M V R K TGTCTCAT ATGTGAGG ATGGAAATG TACTTTTT. CTGAATMAG TACTGTGGA TTTAATATC CCACATTAH TTTCAATAT GGTGTGAG AYDLLS TLAS PSE FSO LTTA RAV ANL AAN LOPK AGGTAAGAT TTATATGA CATTAGCAG CCTTGGAG CDTGTACAT TACATATG: TATCTGTA GATGATGAG CACTCTACT GGTGGAGG PSI LAERYKF RQR RQLN ESIADY LTRLEFL SRH C CCTCAATCT TGGCGAAG AKAAGATTT GCAGAGCAG GCACACTGMA TGTGCTAKA CTGTACTAG TGCAGAAAT AAAAAATA TCAAAAAT EFGSSLDENLRDQHVCGLXSEIIRQRLFAEKKL GTAATFQOG CTGTCTGT GATGAAATC TCGGTGTA ANGTGTTT GAAATMAGA CGAAATCT AAGCAAGG TTATFQOG AAGAAAT EYRRAV TLLSL E AERDAI AVERTP IEEVHEI AGATAZAG CTGTACTTA CTATGCTTA CTGTGGAGA CGATGCTAT TCGATGAGA CTAAGGATA TTAGAGATY TCAAAAAT NFNECSR C G D RRHQRED CIY KDT VCSSCHETGHL AACCTAATG AATGTTGAG ATGTGAGC AGAGACAC AAGCAAGA TGCATATC AAGACTAG TGTGACTC ATGCTAGAA AAGCGATC RRNCPK NGLKHQAEAA GGSARTG ARGHRGAGGH TAGAGAAAT GTGCAGAAA AGCGCTTA AGACAGAGC GAGAGCAGC GGGGFFACG CAGCCAGCG GCGCGGCGG AAGCGGCGG GAGCGGAG ORF 2 RRSMRG QRVR DEEAES GLLC TYYY TIK TTYTHTT QATVARA ASA CRG GREN AFL HLL TDGM HDD TYM H CAGCGATGT TGGCGCGG AGGAGGAG GCGGAGGCG GCGCTCTTG CACTATPAA AGAATCAAA CAGCGGCGC ACATKACA ORF 1 TSHMNR T GVEKNSLCT FHHBQD GCGEHEPHT QHTLSHY EPVCIE VEVO ACTYCAZAA TGAAGAGC GGTGTGAG AGAAAGGA CCGTCTGC GAHAGCAT TACATATA TAAAGCGT TGTATMAG TTAGATGA MCLLRH EVDT GSA LSCISKHYVD ETP SFERLOA AAATGCTG CTACAGAGG AGTGTGAG TGTGTGTA TGTGAGAA TGTTHGAT AAATATTT CTAGAGGA GTTACAGCA CLLHLRFPYDGSII RPLGFIN TIV EYQG VSRHLD L TCTTATTA ATTAAGAT TTAGACTTC TCAATCTC GCTTGTAG GTTATMAG ACTATMAG ACTATMAG TGTTCANAA GTTATAGAT TVIDCG TTH LER QHLARL NINI HSKPT SFKI TGTATGCT AGACAGAA ACAGTATC TATGTGCG KAGATGTA GYRATGA ATATATAT AAAAAATCA AACTAGCA GTTTAAAT QNS HFV TENA RDY NKL IMEI VSR HKS LFDG TIG ACACAGAT ACTTGTAA CCMAGCC: AGAGACT: AMTAAATRA TAAAGCAT TGTCTTGA CTAAGATCT TGTGCGAG CACTTGGT KTY GGTARELI VRF DAVPIY C RAR PVPT ALRER D AAATAGCG GGGZCAGC AGATMAGC GGTGCGCG AGGCGGCC THTCTAGC GCGCGGCC GCGTGTGTA TGTGTGCG: GAGCGCGG AELDAP: LAA GVIF PVDHSD NATP LUV VREAGCG ATYCGAGCT GTGCGCGG GGTGTGCG ACAGTATC CACTGATCT GCGCGGCC ACTYGTGTA GCGCGAGG GCGCGCGG LRI CADYEV T LNE VEA IDRF PVP ENHDLFS HLS TGTGAGAT TGTGAGAT ATAGTATC CTTAAHGA GHTATGGA TGCAGCAT TCGGTGCA AAGATGTA ACTATMAG TACTTGC GHEFTT LDDL SQE VLVQI VLSERS SEYTVINTR C GGATAAAT TTTTACTA CTGATTTA TCTAGTAT AMBAHATF AGCTATMAG CACTCTTA GCGAGAGC GGTATMAG ACAGATG LFXYSR LVTGLAS SPGIFQ RLVNHF RVFVHVU GATTATTA ATATCTCC CTCTTAGC GHTGCTC GHDGCGC ATFTTCAN AACTATGCT AMATGTTT AAAAAATC CAATATGT VTTDDI LRHNQDLDSR LKSIKEV LDILEY GLK AGTFTCTAT GATGAGAT TTGATMAG TGTGCGAG GHTATMAG TAAAGACT TAAAGACT TAAAGACT TAAAGACT TAAAGACT IKRSECF FHV TVRYT LG FII DQH GVRVDPEKVS ATTAAGGA GTATGCGA GTTACAGTA AGGAGMGA GHTATMAG GTTACAGTA GATFAGAG GHTFAGAG GHTFAGAG AAGTCAAT IATNPH PNHVTEL KSP IGH VHY SEF IQDLSAN CAATGAC AATGCGAC CCAAAZAG TGCAGAAAT AAAATCTTC AAGCTAGG TAACTTCTA TGTAGACT ATACAGAT TGTGTACA </pre>	<pre> LSPLYALLREKGRMNMGNKQNAALFLVVRKFLCS 2400 TPTACTCT TTAGTGGC TTTTAAAA AGGAGAGC TGGAGTGG GAGMAGCA AAGCTGTY TGTGAGTG TTAAGATY TTTGTACT 2400 TKALAHFDHSLLESVLTVDASARGLCAVLAQRCP C 2500 ACAAAAGC TGGACATY TGTATGCT TTAGTGGC TTTGACTCT GAGMAGAG GCGGGTGTG TGGYKGGY GTTGGCCAG CCGGCGCAG COERVVATA SRALTTTRELNYSQIREEALAIYFA 2600 GATGTGGA GCGGGTGT GATADUCT CAGCGGCT CACTCTAT GATPTACT ACAGCTAT TCAATMAGA GAATGCGA TGTGTTGC VEFPAQYLTGRKFLRYDNRFLVSIIFGPMI GIP 2700 GOTGAAAA TGTACAT ACTHTERG GAGAAATC AACTGCGA CCGGCGAA ACTGTGTG AGCATTTG GGTACTAT AGGATGCG SAAASRLORMAINLSAYDFEIEYVRYDERVADAL 2800 ACCGGCGAG CTAGGCTT CAGCGCTG TGTGTATG TGTGTATG TGTGTATG TGTGTATG TGTGTATG TGTGTATG TGTGTATG SRLIESQEMDVASEETDLPEOTYLNHSTAEALLI 2900 TPTACTCT AATGACTY CAAAAAATG AAGTACTY: GAGAGAA GACTGTGG AGAAATCA CTACTCT TCAACAGAG CTCTTAT DYMVLKQKTS SDPILS AVLS YLRDGMPLDIEIH 3000 AGATTATY GTCTTAAA ACAAATAG TATGTACA ACTEMAGC GGTATMAG TTATMAG GAGCGAGC GAGCGAGC AGAAATMAG ELKPYTH RHMELY IELG CINHGRVVI PSSCRH K 3100 GAATTAAC CATATTATA TAAAGAAC GAATYATA TGAATMAG AGTATMAG TGTGTATG GGTGTATG GGTGTATG GGTGTATG IITELHDPH HGVETE SLARSYV MFP GID EALE 3200 AAATAAC GAGCTTAT GATGCGATA TGGATAGT GAAACAAA TGTGTAGC GTATMAG TGTGTGCG GATGTAGC GATGCGAG TEC EACTVCAAVA DAPSTRAFRSMPSPSRPMSR 3300 GAGGATGT CAGCTTGA CCGTGTGC TCGTGTGA GCGGCTCT CTRGAGC GCGCGCGC TGGCGCGC GGTGTGTG LHLDFPH PIGGVT VLVVVDSCRMIEA IKHRT Y 3400 TTACTTGT ACTTGTAG CCGATAGT GAGTACTY ACTTGTGT AGGATMAG TGTGTATG GGTGTATG GGTGTATG GGTGTATG AQAVISVLRDLHSEKFLPRTGSDHGFPPSSSD 3500 GCGAGAGC GGTATMAG CACTGCGG ACTYGTGC AAATGCGC CTRGAGC GCGCGCGC GGTGTATG GGTGTATG GGTGTATG FQFLIHNGIKBIYSA THTPASHGAAEHAVRIC 3600 TTTGTAG TTTGTAGT AATGCGAG AATGCGAG THTGAGC CTRGAGC GCGCGCGC GGTGTATG GGTGTATG GGTGTATG KRAIKKALQHLVSTALCR FLLATM TERATT C 3700 AAGTGTGA TAAAGAAC ATTAAGAG AMTAAATG THTGAGC THTGAGC GGTGTATG GGTGTATG GGTGTATG GGTGTATG DSPANILQGRSLRHL DNLKPEROSRVIAQER 3800 CGATMAGC AGCATATA TTAGGCTC GAGCTGCG TGTGTATG GATATMAG AAGCGAGC GATGAGTA GTTGTGCG AAGCGAGC SEQHAGVQR QLEPGT EVMYRDT RGLDMVPGT 3900 GAGAGAGC AAGCGAGC GTGTGCGC AGACTYAG CCGGCGTA AGGATMAG TGTGTATG GGTGTATG GGTGTATG GGTGTATG ILQLCSRDY CVBSDFG THTGAGC THTGAGC AATMAGC GGTGTATG GGTGTATG GGTGTATG GGTGTATG IDKVDTPASIREH ISPELLAQHLDLR TOSRFR 4000 ATATMAGT ACTATMAG TTTGTGTA TGTGTATG GGTGTATG GGTGTATG GGTGTATG GGTGTATG GGTGTATG LSFPHTS GEEPVA VVVDSGKSSSEPPHSTPAQ 4100 CTATGCTT CCGAGTGA GGTGTATG GGTGTATG GGTGTATG GGTGTATG GGTGTATG GGTGTATG GGTGTATG AVS FIRC CAGCTGTG AGGCGCGC TGTGTATG GGTGTATG GGTGTATG GGTGTATG GGTGTATG GGTGTATG KREFPV RIRKFPV RYG FEEIP 4200 CMAAGTGA ACCTGTGA AAGTGTGA AGGCGCGC THTGAGC THTGAGC THTGAGC THTGAGC THTGAGC TACTTGT TGTATG CTGATTA GATATMAG TATATATY AATGCTT ATATMAG GATATMAG GGTGTATG TATGTTTA 4300 AGTATMAG CACTTCTG TGTATMAG CAGTGTAGC AGGCGCGC THTGAGC THTGAGC THTGAGC THTGAGC </pre>
---	--