Volume 12 Number 17 1984 Nucleic Acids Research

Introns and their flanking sequences of *Bombyx mori* rDNA

Haruhiko Fujiwara*, Teru Ogura, Naoko Takada, Naoko Miyajima, Hajime Ishikawa* and Hideaki Maekawa+

Department of Technology, National Institute of Health, Kamiosaki, Shinagawa-ku, Tokyo 141, and *Department of Biology, College of Arts and Sciences, University of Tokyo, Komaba, Meguro-ku, Tokyo 153, Japan

Received 14 June 1984; Revised and Accepted 10 August 1984

ABSTRACT

We obtained two different clones (16 kb and 13 kb) of B. mori rDNA with intron sequence within the 28S-rRNA coding region. The sequence surrounding the intron was found to be highly conserved as indicated in several eukaryotes (Tetrahymena, Drosophila and Xenopus). The 28S rRNA-coding sequence of 16 kb and 13 kb clone was interrupted at precisely the same sites as those where the D. melanogaster rDNA interrupted by the type I and type II intron, respectively. The intron sequences of B. mori were different from those of D. melanogaster. In 16 kb clone, the intron was flanked by 14 bp duplication of the junction sequence, which was also present once within the 28S rRNA-coding region of rDNA without intron. This 14 bp sequence was identical with those surrounding the introns of Dipteran rDNAs.

INTRODUCTION

In the 28S rRNA-coding region of eukaryotic rDNAs are sequences which are evolutionarily conserved (1). This suggests that functionally important roles may be assigned to these regions of the 28S rRNA. It is within one of these conserved regions of rDNA that insertion sequences (introns) are occasionally found (2). One type of intron is present at the position of two third from the 5'-border of the 28S rRNA-coding region (Physarum (type II) (3) and Tetrahymena (4,5) rDNA). In rDNAs of Physarum (type I) (6), mitochondria of yeast (7) and Neurospora (8), and Chlamydomonas chloroplast (9), the other type of intron is present slightly downstream of the former. These introns are co-transcribed with the coding sequences and the transcripts spliced (3-9).

Introns are also known within the 28S rRNA region of Dipteran rDNA. Unlike those described above, the Dipteran rDNA containing introns is not apparently transcribed (10-12), which suggests that these introns are evolutionarily distinct. In case of Drosophila rDNA, both the two classes of intron (type I and type II) have been known (13).

As for Bombyx mori, Manning et al. reported that the repeating unit of

rDNA is uniform and is not with introns (14). However, it is reasonable to question whether the situation is common to the whole species of B. mori since so many distinctive strains are known in this insect. In this communication, we demonstrated presence of introns within the 28S rRNA-coding region of B. mori rDNA. We also located the coding/intron junction by comparison of restriction maps of the intron$^+$ and intron$^-$ rDNA, and subsequently determined the nucleotide sequences surrounding these boundaries.

After the original version of this paper had been submitted for publication, there appeared a paper which demonstrated the presence of intron$^+$ rDNA in a different strain of B. mori (15).

MATERIALS AND METHODS

Cloning of rRNA genes

Recombinant phage library of B. mori genes, strain Gunpo x Shugyoku, was constructed by insertion of partial and complete EcoRI digests (the range of 12 to 21 kb) of 50 to 100 kb DNA from posterior silk glands into the Charon 4A vector (16,17). The library was screened using alkali-digests of rRNA (18) labeled with ^{32}P with polynucleotide kinase or ^{125}I-labeled rRNA as probes (19).

Restriction maps

All the restriction enzymes used were purchased commercially. Appropriate fragments of λclones were subcloned into pBR322 plasmid.

DNA sequencing

Ribosomal RNA-coding/intron boundary regions of each clone was located by comparison of the restriction maps of the three clones. Each fragment containing the boundary region was separated on electrophoresis and sequenced by the method of Maxam and Gilbert (20,21).

RESULTS

Restriction maps of rDNA clones

Three classes of rDNA were screened out of the partial gene library of B. mori (Fig. 1). The first class (λBmR11) was without intron (intron$^-$) and identical to the rDNA repeating unit of 11 kb in pBmR1 constructed by Manning et al. (14). The other two classes were revealed to contain insertion sequence (intron$^+$).

The 16 kb clone (λBmR16) was interrupted at the position of about 4.5 kb inward from the site of EcoRI at the 3'-distal of the 18S rRNA-coding region

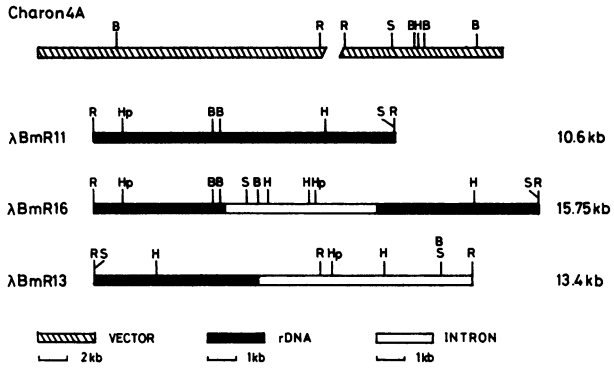

Fig. 1. Comparison of the restriction maps of three different clones of B. mori rDNA. Each clone was screened out of the partial gene library into the Charon 4A vector of B. mori. Ribosomal RNA-coding regions within all the clones were shown in solid blocks and intron regions in open boxes.
B, BamHI; R, EcoRI; Hp, HpaI; S, SmaI; H, HindIII.

(Fig. 1). The size of insertion was 5 kb just like of that in Drosophila melanogaster (type I) and D. virilis rDNA. The other intron$^+$ clone (λBmR13), lacking the 28S rRNA-coding sequence upstream of the intron, contained, at least, two EcoRI sites within the insertion sequence. It is not clear whether the two EcoRI-generated fragments in λBmR13 are colinear in the genomic rDNA. The whole size of the intron corresponding to this clone is not known, either. Yet, this clone no doubt represents another intron$^+$ rDNA in B. mori because the position of intron in it is obviously different from that in the 16 kb clone, which is evidenced below.

To estimate the percentage of rRNA cistrons in B. mori containing introns, the EcoRI-digested genomic DNA blots were probed with λBmR11 (data not shown). By this determination about 85% of the rRNA cistrons were without the intron. About 3 % contained the λBmR16 type. While 3 to 6% were suggested to contain the λBmR13 type, the rest of the cistrons were not characterized detailedly.

These insertion sequences were not hybridized with either rRNA or the non-coding sequence of intron$^-$ rDNA. The two intron sequences were not cross-hybridized, either (Maekawa, unpublished data). Repetitious sequences were found in the 13 kb rDNA intron, while that of the 16 kb clone did not seem to contain such sequences (Ogura et al., in preparation).

The restriction maps of rDNA contained in λBmR11 and λBmR16 were identical with each other along the length from the 5'-end (left-hand) through the 7th

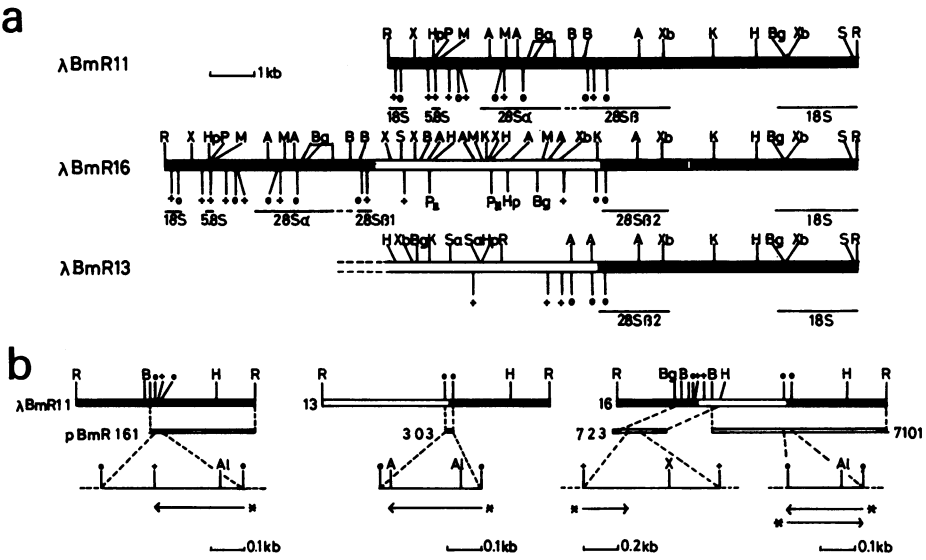

Fig. 2. Detailed restriction maps of three classes of rDNA (a) and sequencing strategies to identify the rDNA intron/coding junctions (b).
(a) Coding regions for 18S, 5.8S and 28S rRNA are shown below each map. Coding regions are indicated as solid blocks and intron regions as open boxes. The small open box in the map of λBmR16 represents the additional sequence less than 50 bp between XbaI and KpnI site (see Text). Restriction sites upstream of the HindIII site within the intron of λBmR13 were not determined.
(b) Subclones prepared from each λphage clone were shown below its map. The subclones were constructed by insertion of each fragment into pBR322 vector. Restriction maps of intron/coding boundary region were shown below each subclone with expanded scale. Astericks and arrows below them represent the start point and direction of DNA sequencing, respectively.
R, EcoRI; X, XhoI; P, PstI; Hp, HpaI; A, AvaI; Bg, BglII; M, MluI; Xb, XbaI; K, KpnI; H, HindIII; B, BamHI; S, SmaI; Sa, SalI; P_{II}, PvuII; Al, AluI; o, PvuI; +, HincII.

HincII site. The 6th PvuI site in the intron$^+$ 16 kb clone was found about 5 kb downstream of that in the intron$^-$ clone. The same PvuI site was also present in the 13 kb intron$^+$ clone, and restriction maps along the length from this PvuI site through the EcoRI site at the 3'-end (right-hand) were common among the three clones, except that an additional sequence of less than 50 bp was present in the region between the XbaI and KpnI site, about 4 kb upstream of the 3'-end of 16 kb clone (Ogura et al., in preparation).

Since the regions of coding/intron boundaries were located, fragments containing these regions were isolated by electrophoresis and sequenced according to a strategy as shown in Fig. 2b.

```
                              -90       -80       -70       -60       -50       -40       -30
B. mori               GCGCGGGTAAACGGCGGGAGTAACTATGACTCTCTTAAGGTAGCCAAATGCCTCGTCATCTAATTAGTGA
D. virilis            .....[                    ]..N...................................G.C.N
D. melanogaster       .........C............................................................
C. erythrocephala     ......................................................................
X. laevis             ......................................................................
T. pigmentosa         ......................................................................

                              -20       -10       +1        10        20        30        40        50
B. m.                 CGCGCATGAATGGATTAACGAGATTCCCACTGTCCCTATCTACTATCTAGCGAAACCACAGCCAAGGGAACGGGCTTGGG
D. v.                 ..........................T....................................................A
D. m.                 ..........................T....................................................A
C. e.                 ..........................T....................................................A
X. l.                 ...............G......................C........................................C
T. p.                 ..................T......A............................C.......T..............A.A

                              60        70        80        90       100       110       120
B. m.                 AGAATCAGCGGGGAAAGAAGACCCTGTTGAGCTTGACTCTAGTCTGGCATTGTAAGGAGACATGAGAGGTGTAGCATAA
D. v.                 .T...T...................T............._.A....A..G....G........A..........A....
D. m.                 .T...T...................T...............A.......G.............A..........A....
C. e.                 .T...T...................................A.......G.............A..........A....
X. l.                 G.............................]
T. p.                 .T....................................._.....AA.T]

                      130       140
B. m.                 GTGGGAGATCGTTTCGCGGC
D. v.                 .........AT.AGACTTCG
D. m.                 .........AT.AGAC.TCG
C. e.                 .........TT.AGTCTTT.
```

Fig. 3. Comparison of the 28S rRNA-coding sequence which flanks the intron. The nucleotide sequence of the 28S rRNA-coding region flanking the intron of B. mori was compared with the equivalent sequences of Drosophila virilis, Drosophila melanogaster, Calliphora erythrocephala, Xenopus laevis and Tetrahymena pigmentosa. Type I intron sequence (in λBmR16) is found in between position -1 and +1 (see Fig. 4).

Sequence conservation of 28S rRNA-coding region of B. mori rDNA

One of the most conserved sequences in the 28S rRNA-coding region of rDNA is that which flanks the introns. The corresponding region of B. mori rDNA (11 kb intron^{-} clone) was sequenced and compared with those of other species. Nucleotides were numbered on the basis of the site of intron. As shown in Fig. 3, homology of the sequence from -99 to +80 of the 28S rRNA-coding region of B. mori to the equivalent region of Tetrahymena, Drosophila and Xenopus rDNA (13,22-24) amounted to 96-98%. Moreover, the B. mori sequence from -29 to +49, except nucleotide -3, was identical with the corresponding region of Dipteran species. The bases at -3, +50, +52, +56, +92, +100, +114 and +125 of B. mori rDNA were invariably different from those of Dipteran rDNA. Smith and Beckingham (25) have pointed out that

Fig. 4. Comparison of the intron/coding boundary sequences of several insect species. The 28S rRNA-coding sequence of rDNA (intron⁻) was shown on the top. Insertion sites for type I and type II intron were indicated by vertical arrows.
(a) Type I intron/coding boundary sequence was determined using DNA fragments from λBmR16 of B. mori (see Fig. 2b), and compared with the equivalent sequence of D. virilis, D. melanogaster (type I intron) and C. erythrocephala. The 14 bp duplicated sequence was underlined. The 8 bp inverted repeat sequence, observed in the region close to the type I insertion site in B. mori and D. melanogaster, was indicated by horizontal arrows. Chi-like sequences within the intron of B. mori rDNA were boxed.
(b) Type II intron/coding boundary sequence was determined using DNA fragments from λBmR13 (see Fig. 2b). The sequence was compared with the corresponding sequence of type II intron of D. melanogaster.

there exists variable region at position between +139 and +160 of Dipteran rDNA. This seems to be true also for B. mori rDNA. At position +139, the B. mori sequence suddenly branched off from those of Diptera (Fig. 3).

Type I and type II introns of B. mori rDNA

Upon sequencing the intron/coding junctions in the 16 kb clone of B. mori, it was indicated that the intron resides at the same position within the 28S rRNA-coding region as do the introns of D. melanogaster (type I) and D.

virilis rDNA (Fig. 4a). Thus, the intron in the 16 kb clone was referred to as type I intron of B. mori. As is the case with the 0.5 kb type I intron of D. melanogaster and D. virilis, this intron of B. mori was surrounded by a duplication of 14 bp. An inverted repeat of 8 bp (5'-ATTCCCAC/GTGGGAAT-3') at the coding region adjacent to the 5'-junction position and the insertion region adjacent to the 3'-one was found specific to the type I intron of B. mori. In the type I intron of D. melanogaster is an inverted repeat sequence of 8 bp (5'-TGTTTTCG/CGAAAAAA-3') at a similar position (Fig. 4a).

The 13 kb clone of B. mori lacked the 5'-side of insertion boundary region. As shown in Fig. 4b, the position of coding/intron boundary in this rDNA was found to be the same as that of the type II intron of D. melanogaster. It was apparent that the boundary sequence of the type II intron is different from that of the type I (Fig. 4a and b).

Apart from the boundary region, both the type I and type II intron sequences of B. mori were different from those of Dipteran rDNA. In an effort to deduce a possible mechanism involved in the introduction of these introns into the rDNA, chi sequence (GCTGGTGG) or the alike was searched for in the boundary region of the introns. As a result, chi-like sequences (GTTGGTGGGA and GGTGGTGGGA) were found within the 3'-distal of the type I intron (Fig. 4a).

DISCUSSION

It has been known that there are evolutionarily conserved sequences within the 28S rRNA-coding region of rDNA. It has been suggested that one of such conserved regions of the rRNA interacts with initiator $tRNA^{met}$ (23,26). This conserved region can be of particular importance in function. In several rDNAs this region is interrupted by insertion sequences of which the transcript is excised in the course of maturation of rRNA. The site of interruption in rDNA seems to be species-specific (1). In B. mori, however, both the type I and type II intron are located at exactly the same position as are those of D. melanogaster, respectively. Moreover, exactly the same 14 bp sequence (TGTCCCTATCTACT) which surrounds both sides of the type I intron of D. virilis (22), C. erythrocephala (25) and D. melanogaster (0.5 kb type I intron) (27,28) was also found to flank the type I intron of B. mori.

Transposability of the intron of Drosophila rDNA is not clear. However, the type I intron appears moderately repeatedly along non-rDNA regions of D.

melanogaster genome and evidence suggests that these copies are actually derived from rDNA (29). Although mobility of the intron of B. mori rDNA is not evidenced, either, it is possible that the type I intron is transposed by means of chi-like sequences. In B. mori rDNA, two chi-like sequences are present with 9 bp interval at the boundary position of the type I intron (Fig. 4). The chi sequence is known as a hot spot concerning DNA recombination of λphage (30). Of late such a hot spot is detected in immunoglobulin gene, too (31). Therefore, it is reasonable to assume that these chi-like sequences of B. mori rDNA, as well as the directly-repeated terminal sequences mentioned above, should play a role in recombination of DNA leading to a transposition of the intron sequence.

Unlike that of D. melanogaster, however, the B. mori type I intron sequence may be unique to the rDNA region in B. mori genome (Ogura et al., in preparation). If this sequence is movable, it is probably looped out of the genomic site and excised to be a circular molecule such as the extrachromosomal circular copia (32,33). So-called att sequence (GCTTTTTTATACTAA) (34) exists on λphage and bacterial genome. This sequence is known to be related to the integration of λphage. The 3'-half (TATACTAA) of att sequence is similar to TATCTAC, 3'-half of B. mori duplicated junction sequence mentioned above. In B. mori rDNA the 8 bp inverted repeat sequences resided upstream adjacent to each of the 14 bp junction sequences. A similar sequence in the type I intron of D. melanogaster was found at positions slightly different from those described above. It has yet to be known whether the intron$^+$ rDNA of B. mori is transcribed.

It is possible that the failure of Manning et al. (14) to detect intron$^+$ rDNA in B. mori is simply due to the strain used by them. Sound evidence for this explanation was provided recently by Lecanidou et al. (15), which demonstrated the presence of intron$^+$ rDNA in B. mori using a different strain. Though the authors did not determine any sequence related to the introns, upon scrutinizing their restriction maps presented it is obvious that they were concerned with the rDNA distinct from ours. This suggests that there is remarkable variance in the rDNA repeats in different strains in B. mori.

ACKNOWLEDGMENTS

We are indebted to Dr. S. Izumi of Tokyo Metropolitan University for screening of the gene library of B. mori. Thanks are also due to Drs. Y. Suzuki, R. Kominami and B. Sakaguchi for their advices and encouragement over the course of this study.

This work was supported in part by grants from the Ministry of Education, Science and Culture and from the Joint Studies Program with Dr. Y. Suzuki of National Institute for Basic Biology.

+To whom correspondence should be addressed.

REFERENCES

1. Gerbi, S.A., Gourse, R.A. and Clark, C.G. (1982) in The Cell Nucleus, Busch, H. and Rothblum, L.I. Eds., Vol. X, pp. 351-386, Academic Press, New York.
2. Beckingham, K. (1982) in The Cell Nucleus, Busch, H. and Rothblum, L.I. Eds., Vol. X, pp. 206-266, Academic Press, New York.
3. Nomiyama, H., Kuhara, S., Kukita, T., Otsuka, T. and Sakaki, Y. (1981) Nucleic Acids Res. 9, 5507-5520.
4. Wild, M.A. (1980) Nature 283, 693-694.
5. Kan, N.C. and Gall, J.G. (1982) Nucleic Acids Res. 9, 2809-2822.
6. Nomiyama, H., Sakaki, Y. and Takagi, Y. (1981) Proc. Natl. Acad. Sci. USA 78, 1376-1380.
7. Dujon, B. (1980) Cell 20, 185-197.
8. Burke, J.M. and RajBhandary, U.L. (1982) Cell 31, 509-520.
9. Allet, B. and Rochaix, J.-D. (1979) Cell 18, 55-60.
10. Alwine, J.C., Kemp, D.J. and Stark, G.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5350-5354.
11. Long, E.O. and Dawid, I.B. (1979) Cell 18, 1185-1196.
12. Jolly, D.J. and Thomas, C.A., Jr. (1980) Nucleic Acids Res. 8, 67-84.
13. Roiha, H., Miller, J.R., Woods, L.C. and Glover, D.M. (1981) Nature 290, 749-753.
14. Manning, R.F., Samols, D.R. and Gage, L.P. (1978) Gene 4, 153-166.
15. Lecanidou, R., Eickbush, T.H. and Kafatos, F.C. (1984) Nucleic Acids Res. 12, 4703-4713.
16. Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connele, C. and Quon, D. (1978) Cell 15, 687-701.
17. Okamoto, H., Ishikawa, E. and Suzuki, Y. (1982) J. Biol. Chem. 257, 15192-15199.
18. Maekawa, H. and Suzuki, Y. (1980) Develop. Biol. 78, 394-406.
19. Benton, W.D. and Davis, R.W. (1977) Science 196, 180-182.
20. Maxam, A.M. and Gilbert, W. (1980) Methods in Enzymol. 65, 499-560.
21. Maniatis, T., Fritsch, E.E. and Sambrook, J. (1982) Molecular Cloning, pp. 254-255, Cold Spring Harbor Lab., New York.
22. Rae, P.M.M., Kohorn, B.D. and Wade, R.P. (1980) Nucleic Acids Res. 8, 3491-3504.
23. Gourse, R.L. and Gerbi, S.A. (1980) Nucleic Acids Res. 8, 3623-3637.
24. Wild, M.A. and Sommer, R. (1980) Nature 283, 693-694.
25. Smith, V.L. and Beckingham, K. (1984) Nucleic Acids Res. 12, 1707-1724.
26. Dahlberg, J.E., Kintner, C. and Lund, E. (1978) Proc. Natl. Acad. Sci. USA 75, 1071-1075.
27. Roiha, H. and Glover, D.M. (1981) Nucleic Acids Res. 9, 5521-5532.
28. Dawid, I.B. and Rebbert, M.L. (1981) Nucleic Acids Res. 9, 5011-5020.
29. Kidd, S.J. and Glover, D.M. (1980) Cell 19, 103-119.
30. Smith, G.R., Kunes, S.M., Schultz, D.W., Taylor, A. and Triman, K.L. (1981) Cell 24, 429-436.
31. Kenter, A.L. and Birshtein, B.K. (1981) Nature 293, 402-404.
32. Flavell, A.J. and Ish-Horowicz, D. (1981) Nature 292, 591-595.
33. Flavell, A.J. and Ish-Horowicz, D. (1983) Cell 34, 415-419.
34. Landy, A. and Ross, W. (1977) Science 197, 1147-1160.