Novel non-autonomous transposable elements on W chromosome of the silkworm, *Bombyx mori*

HIROAKI ABE¹, TSUGURU FUJII², TORU SHIMADA² and KAZUEI MITA³

¹Department of Biological Production, Faculty of Agriculture, Tokyo University of Agriculture and Technology, Saiwai-cho 3-5-8, Fuchu, Tokyo 183-8509, Japan
²Department of Agricultural and Environmental Biology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113-8657, Japan
³National Institute of Agrobiological Science, Owashi 1-2, Tsukuba, Ibaraki 305-8634, Japan

Abstract

The sex chromosomes of the silkworm *Bombyx mori* are designated ZW (XY) for females and ZZ (XX) for males. Numerous long terminal repeat (LTR) and non-LTR retrotransposons, retroposons and DNA transposons have accumulated as strata on the W chromosome. However, there are nucleotide sequences that do not show the characteristics of typical transposable elements on the W chromosome. To analyse these uncharacterized nucleotide sequences on the W chromosome, we used whole-genome shotgun (WGS) data and assembled data that was obtained using male genome DNA. Through these analyses, we found that almost all of these uncharacterized sequences were non-autonomous transposable elements that do not fit into the conventional classification. It is notable that some of these transposable elements contained the *Bombyx* short interspersed element (Bm1) sequences in the elements. We designated them as secondary-Bm1 transposable elements (SBTEs). Because putative ancestral SBTE nucleotide sequences without Bm1 do not occur in the WGS data, we suggest that the Bm1 sequences of SBTEs are not carried on each element merely as a package but are components of each element. Therefore, we confirmed that SBTEs should be classified as a new group of transposable elements.

Introduction

The sex chromosomes of the silkworm *Bombyx mori* are designated ZW (XY) for females and ZZ (XX) for males (Tanaka 1916). The W chromosome is interesting from the viewpoint of sex determination. Development as a female is determined by the presence of a single W chromosome, irrespective of the number of autosomes or the Z chromosome (Hashimoto 1933). Therefore, it is presumed that the female-determining gene (*Fem*) is present on the W chromosome. Interestingly, the W chromosome does not recombine with the Z chromosome or autosomes and it remains static without undergoing any changes during mother-to-daughter transmission. Another interesting feature of the W chromosome is that although 400 or more visible mutations have been placed on the linkage map of *B. mori* (Doira 1992; Goldsmith 1995; Goldsmith et al. 2005) and there is evidence for the presence of numerous genes on the Z chromosome (Fujii et al. 1998; Koike et al. 2003), the W chromosome is devoid of functional genes except for the putative Fem gene. No W-specific markers have been found, despite the construction of extensive linkage maps for the silkworm, which includes random amplified polymorphic DNA (RAPD) (Promboon et al. 1995; Yasukochi 1998), amplified restriction fragment polymorphisms (Tan et al. 2001), SSRs (Miao et al. 2005), and SNPs (Yamamoto et al. 2006). A draft sequence of the *B. mori* (p50 strain) genome using 3-fold and 5.9-fold whole-genome shotgun (WGS) sequencing was obtained by Mita et al. (2004) and Xia et al. (2004). Recently, these two datasets were combined with

Keywords. W chromosome; transposon; retrotransposon; SINE; sex chromosome; *Bombyx mori*.
the newly obtained fosmid and bacterial artificial chromosome (BAC) end sequences as Build2 data (The International Silkworm Genome Consortium 2008). However, these shotgun sequencing efforts were undertaken using only the male genome. Thus, a systematic molecular analysis of the W chromosome of B. mori has not yet been performed. However, the W chromosome has so far been analysed separately from WGS sequencing. To date, we have identified 12 W-specific RAPD markers in B. mori and one in B. mandarinina (Abe et al. 1998a, 2003). In addition, we obtained two lambda phage clones and found that these DNA sequences comprised a nested structure of numerous retroposons (Abe et al. 1998b, 2000, 2002; Ohbayashi et al. 1998). Further, we obtained 14 W-specific BAC clones and subjected them to shotgun sequencing. Through these analyses, we found that numerous long terminal repeat (LTR) and non-LTR retrotransposons, retroposons, DNA transposons, and their derivatives have accumulated on the W chromosome as strata (Abe et al. 2000, 2002, 2005a, 2008). However, we found that the nucleotide sequences do not show the characteristics of typical transposable elements (Abe et al. 2005a). Although it has been predicted that almost all transposable elements on the autosomes and the Z chromosome have degenerated because of recombination or rearrangement, genome studies in B. mori using male genome DNA would allow us to examine numerous fragments of transposable elements and repetitive sequences. Therefore, we used the WGS and Build2 data to analyse the uncharacterized nucleotide sequences of the W chromosome. These results indicate that almost all of these uncharacterized sequences are non-autonomous transposable elements that do not fit into the conventional classification. Surprisingly, some of these transposable elements contain the Bm1 sequences. The Bm1 element is a major short interspersed element (SINE) in the genome of B. mori (Adams et al. 1986). Therefore, the Bm1 element is believed to be analogous to the human Alu element. Numerous SINEs have been characterized to date (Oshima and Okada 2005). However, to our knowledge, in the case of Alu elements, there is no report of a transposable element containing an Alu sequence (International Human Genome Sequencing Consortium 2001). For this reason, we focus our present analyses on transposable elements containing Bm1 sequences and report their characteristics. In addition, we report the non-autonomous transposable elements identified during the analysis of the W chromosome.

**Experimental procedures**

A B. mori BAC library, designated as RPCI-96, was constructed using genomic DNA extracted from a mixed population of the silkworm strain p50. The library contained approximately 36,000 clones with an average insert size of 168 kb (Koike et al. 2003). Two other B. mori BAC libraries, designated as p50-library and C108-library, were also constructed using genomic DNA of the p50 and C108 strains, respectively (Wu et al. 1999). We used PCR to obtain BAC clones that contained W-specific RAPD marker sequences. Thus far, we have obtained 14 W-specific BAC clones and we subjected these to shotgun sequencing. For this study, we selected the BAC clones 45123 (W-Kabuki RAPD), 5A2G (W-Musashi RAPD), 544H03 (W-Rikishi RAPD) and 522N19 (W-Rikishi RAPD).

We estimated the copy number of each non-autonomous transposable element by conducting a BLASTN search using Build2 data (The International Silkworm Genome Consortium 2008). When using the 5’ or 3’ 200-bp sequence of each element as a query, we used hit numbers showing an E value of more than e – 20 as copy numbers.

To prepare a standard Bm1 sequence for comparison with the Bm1 sequences in the non-autonomous transposable elements, we selected a Bm1 (444 bp) inserted into the Kabuki retrotransposon on the W chromosome because this Bm1 is believed to change at a relatively slow rate (Abe et al. 2000).

**Accession numbers**

The nucleotide sequence data will appear in DDBJ, EMBL, and GenBank nucleotide sequence databases under the accession nos: AB455913 (ChoBm1), AB455915 (Neet), AB455920 (BmImodoki), AB455942 (Tenshi), AB455934 (Hikikomori), AB455926 (Ins11), AB480246 (Rinne-L), AB455911 (Rinne-S), AB455947 (Yaoho), AB455939 (Nagisa), AB455921 (Sayama), AB455940 (Kohichi), AB455953 (Kusanagi), AB455933 (Chikuri), AB455932 (Minichikuri), AB455944 (Shirigonomi), AB455931 (Amanoya), AB455935 (Supermite-BMC1), AB480233 (W-5A2G-C02), AB480234 (W-5A2G-C03), AB480235 (W-5A2G-C09), AB480236 (W-5A2G-C15), AB480237 (W-5A2G-C16), AB480238 (W-5A2G-C17), AB480239 (W-5A2G-C19), AB480240 (W-1K7D-C16), AB480241 (W-544H03-C25), AB480242 (W-544H03-C26), AB480243 (W-522N19-C6), AB480244 (W-522N19-C7), and AB480245 (W-522N19-9).

**Results**

**Secondary-Bm1 transposable elements (SBTEs)**

We analysed the nucleotide sequences of the W chromosome that did not show the characteristics of typical transposable elements. These sequences were repetitive. We think that these repetitive sequences are non-autonomous transposable elements because they do not have a coding region or a degenerate coding region. Moreover, some of these repetitive sequences contain the Bm1 sequence. We analysed the WGS data to find out if there were ancestral states of these sequences that did not contain the Bm1 sequence (figure 2). However, the ancestral states were not found in the WGS data. Based on the nucleotide sequences, length homogeneity of each element, and the copy numbers in the genome, we suggest that these elements containing the Bm1 sequence are not fortuitously inserted pseudogenes but are
transposable elements. We designated these transposable elements as secondary-Bm1 transposable elements (SBTEs) and have described their characteristics below.

**ChoBm1:** The BAC clone 45J23S containing the W-Kabuki RAPD marker sequence has been analysed previously (Abe et al. 2005a). Stretches of 45J23S contained the region that included the Bm1-element sequence, which could not be characterized at that time. The BLASTN search using the WGS data revealed that the sequence containing the Bm1 element is a repetitive sequence (nucleotide positions 1727–7153 in accession number AB126050). We designated this repetitive sequence as *ChoBm1* (figure 1A). Based on the sequence in AB126050 with slight modification according to the WGS data, we determined the consensus sequence of *ChoBm1* (5412 bp). The Bm1 sequence in *ChoBm1* occupied nucleotide positions 3187–3432 corresponding to nucleotide positions 196–442 of standard Bm1. We estimated the copy number of *ChoBm1* per haploid genome to be approximately 60; however, we could not establish the preferential insertion sequence of *ChoBm1*. *ChoBm1* did not contain any long open reading frames (ORFs), and deduced amino acid sequences did not show similarity to previously reported transposable elements. Therefore, *ChoBm1* may be a non-autonomous transposable element. Initially, we thought that the Bm1 sequence was inserted fortuitously into this position on *ChoBm1*. If this were so, the putative ancestral *ChoBm1* that does not contain the Bm1 sequence, as shown in figure 2, would be found in the WGS data. However, we could not find the ancestral *ChoBm1* sequence in the WGS data. Therefore, we believe that the structure of *ChoBm1* shown in figure 1A is the unit of transposition.

![Figure 1. Schematic diagrams of SBTEs. A, ChoBm1 (5412 bp); B, Neet (4046 bp); C, Bm1modoki (1178 bp); D, Tenshi (1360 bp); E, Hikikomori (1740 bp); F, Ins11 (1088 bp); G, Yaocho (5851 bp); H, Rinne-L (3695 bp); I, Rinne-S (3437 bp); PBS, primer-binding site.](image-url)
**Neet:** The W-5A2G-C16 (15,291 bp) and W-5A2G-C15 (6343 bp) stretches of the BAC clone 5A2G contained numerous repetitive sequences. The BLASTN and BLASTX searches using WGS and GenBank data revealed that there was a non-LTR retrotransposon divided between the two stretches because of the sequence gap. We designated this non-LTR retrotransposon as *Taguchi*. *Taguchi* was inserted into the LTR retrotransposon designated as *Yokohama* with target site duplication (5'-AGAGTAAAAAAGC-3'). Moreover, *Yokohama* was inserted into the repetitive sequence (nucleotide positions 614–3326 in W-5A2G-C15 and 10,525–11,862 in W-5A2G-C16) with target site duplication (5'-CATAG), as shown in figure 3. We designated this repetitive sequence as *Neet* (4046 bp) (figure 1B). The Bm1 sequence in *Neet* occupied nucleotide positions 2296–2450 corresponding to nucleotide positions 261–415 of standard Bm1. We estimated the copy number of *Neet* per haploid genome to be approximately 70; however, we could not establish the preferential insertion sequence of *Neet*.

*Neet* did not contain any long ORFs; therefore, it is believed to be a non-autonomous transposable element. Similar to the case of *ChoBm1*, we could not find the putative ancestral *Neet* that did not contain the Bm1 sequence in the WGS data. Therefore, we think that the structure of *Neet* shown in figure 1B is the unit of transposition.

**Bm1modoki:** The W-5A2G-C16 stretch (15,291 bp) contained numerous repetitive sequences. The BLASTN search using the WGS data revealed that the repetitive sequence containing the Bm1 sequence (nucleotide positions 13,667–14,630) was inserted into another repetitive sequence *Seppuku* (data not shown). We designated this repetitive sequence as *Bm1modoki* (1178 bp) (figure 1C). The Bm1 sequence in *Bm1modoki* was at nucleotide positions 337–478 corresponding to nucleotide positions 302–439 of standard
Silkworm sex chromosome transposable elements

Bm1. The estimated copy number of Bm1modoki was approximately 450 per haploid genome. Bm1modoki was inserted preferentially into the 5′-AAA-Bm1modoki-3′ sequence site:

5′-AAA-Bm1modoki-TTT-3′.

Bm1modoki did not contain any long ORFs. Therefore, Bm1modoki is believed to be a non-autonomous transposable element. Based on the copy numbers in the genome, Bm1modoki is believed to be a retrotransposable element. Similar to the cases of ChoBm1 and Neet, we could not find the putative ancestral Bm1modoki sequence in the WGS data. Therefore, we think that the structure of Bm1modoki shown in figure 1C is the unit of transposition.

Tenshi: The W-5A2G-C17 stretch (6003 bp) contained numerous repetitive sequences. The BLASTN and BLASTX searches using WGS and GenBank data revealed that there was a non-LTR retrotransposon with the (GATTT)\(n\) tail. We designated this non-LTR retrotransposon as Takuya. There were several copies of Takuya in the W-chromosome BAC clones. We compared the (GATTT)\(n\) tails of these copies. As shown in figure 4, although there were slight differences in the number of Ts and As between copies, the (GATTT)\(n\) tails had approximately the same structure. We found a sequence containing three Bm1 sequences in the (GATTT)\(n\) tail of Takuya of the W-5A2G-C17 stretch. We designated this sequence as Tenshi (1360 bp) (figure 1D). Tenshi contained three Bm1 element sequences (Bm1 (a), (b), and (c); nucleotide positions 125–362, 402–572, and 1160–1216, respectively). These positions correspond to nucleotide positions 439–196, 272–441, and 287–231 of standard Bm1, respectively. We estimated the copy number of Tenshi to be one per haploid male genome. In addition to the copy on the W chromosome, another copy of Tenshi was found on chromosome 7. Thus, there was one Hikikomori each on the W chromosome and on chromosome 7. Although the copy number of Hikikomori was small, we think that Hikikomori is a transposable element and the structure shown in figure 1E is the unit of transposition.

Yaocho: The BLASTN search using the WGS data showed an LTR-type repetitive sequence containing the Bm1 sequence extending from the W-5A2G-C16 (15291 bp), -C19 (1250 bp), and -C02 (4795 bp) stretches to the W-5A2G-C03 (18,645 bp) stretch (data not shown). We designated this repetitive sequence as Yaocho. A non-LTR retrotransposon, BMC1, was inserted into Yaocho. By eliminating BMC1, based on the WGS data, we determined the consensus sequence of Yaocho (5851 bp; figure 1G). The 5′-LTRs and 3′-LTRs were each 627 bp in length. Although long ORFs were not obtained, the amino acid sequences deduced from nucleotides 2280–2950 of Yaocho showed similarity to the RNase H of micropia of D. melanogaster (Lankenau et al. 1988). Yaocho contained a 5′-truncated Bm1 element sequence (nucleotide positions 4139–4377). Moreover, Yaocho contained a repeat

Hikikomori: In the W-1K7D-C16 stretch (14,359 bp), the BLASTN search using the WGS data revealed a repetitive sequence containing the Bm1 sequence (nucleotide positions 6301–7349). Moreover, this repetitive sequence extended from the W-1K7D-C16 to W-1K7D-C17 because of a copia-like retrotransposon, the LTR retrotransposon Kabuki, and Bm1 insertions (data not shown). We designated this repetitive sequence as Hikikomori (1740 bp) (figure 1E). Hikikomori contained two Bm1 element sequences (Bm1 (a) and (b); nucleotide positions 490–791 and 1049–1292). These positions correspond to nucleotide positions 142–443 and 196–440 of standard Bm1, respectively. We estimated the copy number of Hikikomori to be one per haploid male genome. In addition to the copy on the W chromosome, another copy of Hikikomori was found on chromosome 7. Thus, there was one Hikikomori each on the W chromosome and on chromosome 7. Although the copy number of Hikikomori was small, we think that Hikikomori is a transposable element and the structure shown in figure 1E is the unit of transposition.

Figure 4. The comparison of the (GATTT)\(n\) tails of four copies of non-LTR retrotransposon Takuya. 1C7C-C07, 5A2G-C09, 522N19-C09 and 5A2G-C17 are W-BAC clone names and stretch numbers. Tenshi was inserted in the tail of Takuya of W-5A2G-C17.

Journal of Genetics, Vol. 89, No. 3, September 2010 379
sequence, which we designated as Makabe (nucleotide positions 4499–4899). It included nearly identical 135-bp LTR-like sequences at both ends. Two tandem repeat sequences were found in the internal region of Makabe; no ORFs were identified, despite the presence of the LTR-like sequences. The BLASTN search confirmed that the Makabe is a part of Yaocho. Using the nucleotide sequence from 541 to 740 of Yaocho, we estimated the copy number of Yaocho per haploid genome to be approximately 5.

We initially suspected that Yaocho on the W-5A2G BAC clone degenerated after insertion, and then Bm1 and Makabe were inserted into this element, respectively. Therefore, we thought that ancestral Yaocho copies, which were complete and did not contain the Bm1 sequence, would be found in the WGS data. However, the sequences of the Yaocho parts in the WGS data contained the Bm1 and Makabe sequences as the copy in the W-5A2G BAC clone. Therefore, although long ORFs were not present and an ancestral copy was not found, we consider the structure of Yaocho shown in figure 1G to be the unit of transposition.

**Rinne-L and Rinne-S:** A BLAST search using the WGS data showed an LTR-type repetitive sequence containing the Bm1 sequence (nucleotide positions 9391–13085) in the W-544H03-C26 stretch (20,035 bp). We designated this repetitive sequence as Rinne-L (3695 bp) (figure 1H). A nearly identical sequence was found in the W-5A2G-C09 stretch (32,287 bp). We designated this sequence (nucleotide positions 4–3427) as Rinne-S. Based on the WGS data, adding 13 bp at the 3′ end, we determined the consensus sequence of Rinne-S (3437 bp) (figure 11).

The 5′-LTRs and 3′-LTRs of Rinne-L were each 365 bp in length. A region adjacent to the 5′-LTR was nearly identical to the RNA1em primer-binding site (PBS) of micropia of D. melanogaster (Lankenau et al. 1988). Therefore, this region (5′-TCAGAAGTGAGAT-3′) seems likely to be a putative PBS of Rinne-L. However, we could not find a typical polypurine tract in the region adjacent to the 3′-LTR. Although the deduced amino acid sequence between the 5′-LTR and the Bm1 sequence was similar to micropia of D. melanogaster, long ORFs were not found. Rinne-L contained three Bm1 element sequences (nucleotide positions 944–1023, 1057–1111, and 1419–1608). These positions correspond to nucleotide positions 1–79, 103–157, and 233–422 of standard Bm1, respectively. Two tandem repeats were found in Rinne-L from nucleotide position 2830 to position 2955.

In Rinne-S, the 5′-LTRs and 3′-LTRs were each 375 bp in length, slightly longer than those of Rinne-L, and a region corresponding to nucleotide positions 1302–1755 of Rinne-L was deleted, as shown in figures 1, H&I. There were three tandem repeats in Rinne-S.

Unfortunately, in the Build2 data a long retrotransposable element, such as the full-length BMC1 (Abe et al. 1999b), is converted to a long ‘N’ stretch except for the several hundred base pairs at 5′ and 3′. It was, therefore, very difficult to estimate the copy numbers of Rinne-L and Rinne-S separately. Hence, we estimated the total copy number of Rinne-L and Rinne-S per haploid genome; it was approximately 250.

Similar in case of Yaocho, we suspected that Rinne-L in the W-544H03-C26 stretch and Rinne-S in the W-5A2G-C09 stretch degenerated after insertion into each position, and then the Bm1, TREST, and tandem repeat were inserted into these positions, respectively. However, the sequences of the parts of Rinne-L and Rinne-S in the WGS data contained the Bm1, TREST, and tandem repeat sequences similar to the copies in W-544H03-C26 and W-5A2G-C09. Therefore, although long ORFs are not contained or an ancestral copy is not found, we consider that the structures of Rinne-L and Rinne-S shown in figure 1, H&I are the units of transposition, respectively.

**Ins11:** Ins11 (1088 bp) was previously identified as the putative insertion sequence in the intergenic sequence between the two alkaline phosphatase genes (Itoh et al. 2003). However, at the time, this Ins11 could not be analysed in detail because the WGS and Build2 data had not been completed. Here, we can classify Ins11 as an SBTE. Unfortunately, Ins11 on the nucleotide sequence data of the W chromosome was divided into stretches (details not shown). Therefore, we selected Ins11 from the intergenic sequence between the two alkaline phosphatase genes (Itoh et al. 2003) as the standard copy of Ins11. The Bm1 sequence in Ins11 was at nucleotide positions 325–756 corresponding to nucleotide positions 2–439 of standard Bm1. The estimated copy number of Ins11 was approximately 400 per haploid genome. Although Ins11 in the intergenic sequence between the two alkaline phosphatase genes was inserted into the position as 5′-CTGTAA-Ins11-CTGTAA-3′, almost all copies of Ins11 in the genome were inserted preferentially as 5′-NNNTTA-Ins11-CTGTAA-3′.

Ins11 did not contain any long ORFs. Based on the copy number in the genome, Ins11 is believed to be a non-autonomous retrotransposable element. Similar to the cases of other SBTEs, the putative ancestral Ins11 sequence (not containing the Bm1 sequence) could not be found in the WGS data. Therefore, we think that the structure of Ins11 shown in figure 1F is the unit of transposition.

**Alignment of Bm1 sequences contained in SBTEs**

We aligned the nucleotide sequences of the Bm1 domain of SBTEs (figure 5). Although the length of Bm1 sequences varied, the 3′nucleotide sequences (nucleotide positions from 200 to the poly(A) tail of standard Bm1) are highly conserved in these SBTEs. In particular, the Bm1 sequence of Ins11 was 91% identical (406/444) to the nucleotide sequence of standard Bm1.

**Non-autonomous transposable elements except SBTEs**

We analysed the nucleotide sequences except for SBTEs. The W-522N19-C6 (29,971 bp) and W-522N19-C7 (25,948 bp)
stretches contained numerous transposable elements, especially non-LTR retrotransposons. However, we found that there were nucleotide sequences not showing the characteristics of typical transposable elements as shown in figure 6. The BLASTN search using the WGS and Build2 data revealed that these sequences are repetitive. We designated these repetitive sequences as Kohichi, Nagisa, Sayama, and Kusunagi, respectively (figure 6). These four repetitive sequences did not contain any long ORFs, and deduced amino acid sequences did not show similarity to previously reported transposable elements. In addition, we described the characteristics of the nucleotide sequences designated as Shirigonomi, Chikuri, Minichikuri and Supermite-BMC1.

**Nagisa:** Nagisa (390 bp) was inserted into another repetitive sequence Kohichi (figures 6 and 7A). The estimated copy number of Nagisa was approximately 700 per haploid genome. Nagisa was inserted preferentially into the (TA)n
Figure 6. Schematic diagram of the W-522N19-C6 and partial W-522N19-C7. This map is based on DNA sequence information. Each block indicates a transposable element. The arrows under or over each box indicate the transcriptional orientation. The non-LTR BMC1 was divided into W-522N19-C6 and W-522N19-C7 stretches. Two Kohichi copies were recognized in the non-LTR retrotransposon Kimutaku. One copy contained Nagisa, Bmmariner, and Sayama, but another copy was deleted by insertion of the LTR retrotransposon Yokohama. The left end of the W-533N19-C6 stretch was the ORF region of this Yokohama.

Figure 7. Schematic diagrams of the non-autonomous transposable elements. A, Nagisa (390 bp); B, Sayama (952 bp); C, Kohichi (1302 bp). Boxes with arrow at both ends indicate terminal inverted repeat; D, Kusanagi (421 bp). (TGA)n, TGATGATGA tail. E, Chikuri (712 bp). (A)n, poly(A) tail. F, Minichikuri (413 bp). The region surrounded with dotted lines indicates the deleted region from Chikuri; G, Shirigonomi (279 bp); H, Amanoya (834 bp); I, Supermite-BMC1 (1991 bp).

(n > 2) sequence site. Numerous patterns of insertion such as 5′-(TA)n-Nagisa-(TA)n or 5′(TA)n-Nagisa-(TA)n were found in the genome. There was no difference in copy numbers when using the 5′ 200-bp sequence or the 3′ 200-bp sequence as a query. Moreover, Nagisa did not have a poly(A)-like tail. Therefore, it is believed that the insertion mechanism of Nagisa is different from that of Bm1.

Sayama: Sayama (952 bp) was inserted into another repetitive sequence Kohichi (figures 6 and 7B). The estimated copy number of Sayama was approximately 50 per haploid genome. There was no difference in the estimated copy numbers when using the 5′ 200-bp sequence or the 3′ 200-bp sequence as a query. Moreover, Sayama did not have a poly(A)-like tail. Therefore, it is believed that the insertion mechanism of Sayama is different from that of Bm1.

Kohichi: Kohichi (1302 bp) appeared by eliminating the inserted sequences of Nagisa, Bmmariner (Robertson and Walden 2003), and Sayama. Kohichi was inserted into the non-LTR retrotransposon Kimutaku (figures 6 and 7C). The estimated copy number of Kohichi was approximately 540 per haploid genome. There was no difference in the copy numbers when using the 5′ 200-bp sequence or the 3′ 200-bp sequence as a query. Kohichi was inserted preferentially into the 5′-TACTAC-3′ site: 5′-TAC-Kohichi-TAC-3′. Kohichi has an incomplete terminal in-
verted repeat (TIR) at both the 5′- and 3′-ends: 5′-TIR is 5′-TACCGTAAATGGGCGATTGAGGA-3′, and 3′-TIR is 5′-TCCCTTAATC CCCCATTTTACCCGTA-3′. This incomplete TIR is a characteristic of Kohichi in the genome.

**Kusanagi:** Kusanagi (421 bp) was inserted into the non-LTR retrotransposon Kimutaku (figure 6). Kusanagi had the TGATGATGA sequence at one end (7D). We thought that this (TGA)₃ was the 3′ end of Kusanagi. There was a difference in the estimated copy numbers when using the 5′ 200-bp sequence or the 3′ 200-bp sequence as a query. The estimated copy number when using the 5′ 200-bp sequence was approximately 150 per haploid. However, when using the 3′ 200-bp sequence, the estimated copy number was approximately 600. This is probably due to the 5′-truncation. We think that Kusanagi transposes by retrotransposition, similar to Bm1, with a poly(A) tail. The 5′-truncations of Kusanagi probably result from premature termination of reverse transcription.

**Chikuri:** Chikuri (712 bp) was inserted into another repetitive sequence Kasu (figure 3). Chikuri had the poly(A) sequence at one end (figure 7E). We think that this poly(A) is the 3′ end of Chikuri. Chikuri was preferentially inserted into the poly(A) sequence as (A)n-Chikuri-(A)n. Therefore, the precise boundary between the poly(A) tail of Chikuri and the poly(A) of the harbour site could not be determined. There was a difference in the estimated copy numbers when using the 5′ 200-bp sequence or the 3′ 200-bp sequence as a query. The estimated copy number when using the 5′ 200-bp sequence was approximately 350 per haploid. When using the 3′ 200-bp sequence, the estimated copy number was approximately 1600. Therefore, we think that Chikuri transposes by retrotransposition similar to Bm1. The 5′-truncations of Chikuri probably result from premature termination of reverse transcription.

**Minichikuri:** Minichikuri (413 bp) was identified during the analyses of Chikuri. Minichikuri is believed to be an internal deletion (293 bp) type of Chikuri, as shown in figure 7F. We suspected that Chikuri was generated by the insertion of the sequence that was deleted in Minichikuri. However, the putative deleted sequence from Chikuri was not found alone in the WGS data. The estimation of copy number of Minichikuri is difficult because it has a sequence in common with Chikuri. However, the estimated copy number when using the 5′ 200-bp sequence was approximately 1700 per haploid. Therefore, Minichikuri has more copy numbers and transposed independently from Chikuri.

**Shirigonomi:** Shirigonomi (279 bp) was found in another repetitive sequence in the W-544H03-C25 stretch (data not shown). Shirigonomi was preferentially inserted into the 3′ end of the poly(A) sequence as poly(A)-Shirigonomi. Shirigonomi itself did not have a poly(A) tail (figure 7G). The estimated copy number of Shirigonomi was approximately 40 per haploid genome. There was no difference in the estimated copy numbers when using the 5′ 200-bp sequence or the 3′ 200-bp sequence as a query. Therefore, it is believed that the insertion mechanism of Shirigonomi is different from that of Bm1.

**Amanoya:** In the W-522N19-C9 stretch (34,087 bp), the BLASTN search using the WGS data revealed a repetitive sequence (nucleotide positions 23,631–24,464; data not shown). We designated this repetitive sequence as Amanoya (834 bp; figure 7H). The estimated copy number of Amanoya was approximately 160 per haploid genome. We could not establish the preferential insertion sequence of Amanoya. There was no difference in the estimated copy numbers when using the 5′ 200-bp sequence or the 3′ 200-bp sequence as a query. Amanoya appears not to have a 3′ tail. Therefore, it is believed that the insertion mechanism of Amanoya is different from that of Bm1.

**Supermite-BMC1:** Previously, we found an amino acid coding sequence that is a part of the non-LTR retrotransposon BMC1 in the 45J23 BAC clone (accession no. AB126052; Abe et al. 2005a). Moreover, this coding sequence was flanked with a long IR (500 bp), as shown in figure 7I. In this study, we analysed this sequence in detail using the WGS data. Although one frameshift was recognized, the deduced amino acid sequence was 89% identical (248/279) to part of the amino acid sequence of the endonuclease domain in BMC1 (figure 8). This result strongly indicates that the region between long IRs was derived from BMC1. We designated this sequence (nucleotide position 13270–15260 in AB126052) as Supermite-BMC1. Initially, we thought that Supermite-BMC1 was the result of the insertion of 5′-truncated BMC1 into the DNA transposon-like sequence with the subsequent deletion of the 3′ region of the BMC1 (reverse transcriptase coding region with poly(A) tail). If this were so, the putative ancestral DNA transposon-like sequence would be found in the WGS and Build2 data. However, we could not find an ancestral DNA transposon-like sequence that did not contain the endonuclease coding sequence of BMC1. Moreover, the long IR sequences in the WGS data were flanked with the endonuclease coding sequence of BMC1. Therefore, we consider that Supermite-BMC1 is the transposable element.

It was very difficult to estimate copy numbers because almost the entire Supermite-BMC1 region is occupied with long IR and endonuclease coding sequences. Therefore, we used the sequences between the 5′ IR and the endonuclease domain, as well as between the 3′ IR and the endonuclease domain (nucleotide positions 412–611 and 1312–1511, respectively) as a query. The estimated copy number of Supermite-BMC1 was approximately 50 per haploid genome.

**Discussion**

Although typical LTR and non-LTR retrotransposons, Bm1 (retroposon), and DNA transposons have accumulated as
strata on the W chromosome (Abe et al. 2000, 2002, 2005a,b), there are long and short nucleotide sequences that do not show the characteristics of typical transposable elements (Abe et al. 2005a). Until recently, we could not characterize these nucleotide sequences because there was no large-scale genome analysis data. However, we could use the WGS and Build2 data to analyse the uncharacterized nucleotide sequences on the W chromosome. The results in this study show that almost all of these nucleotide sequences are SINEs, miniature inverted-repeat transposable element (MITE)-like sequences, and transposable elements that do not fit into the conventional classification.

**Secondary-Bm1 transposable elements (SBTEs)**

Bm1 elements were identified originally as SINEs in the genome of B. mori (Adams et al. 1986). It has been estimated that there is a minimum of 2.3 × 10^4 copies of Bm1 per haploid genome (Adams et al. 1986). Therefore, Bm1 elements are generally believed to be analogous to Alu elements in primates. In humans, only a few Alu elements seem to be retrotransposition-competent as ‘master’ genes. Individual Alu copies contain an internal RNA polymerase III promoter but this promoter is not sufficient for active transcription in vivo because appropriate flanking sequences are required for its activation. Therefore, most Alu copies in the human genome are non-functional relics. The probability of two independent Alu insertions occurring in the same genetic region in the human population is essentially zero. The chromosome regions that share the Alu insertion at the same position have inherited it from a common ancestor (reviewed by Batzer and Deininger 2002). We have considered the characteristics of Alu elements in humans to be the same as those of Bm1 elements in B. mori. Therefore, when the same sequences containing Bm1 elements on the W chromosome were found in the WGS data of males, we suspected that these sequences resulted from contamination of male genomic DNA with a minute amount of female DNA during WGS analyses. If this were so, we would have been able to find the W chromosome-specific sequences (Abe et al. 2008, 2005a,b) in the WGS data. However, no W-specific sequences, for example, the nested structure of retrotransposable elements, were found in the WGS data. Therefore, we consider SBTEs to be transposable elements. Recently, a unified classification system for eukaryotic transposable elements was proposed (Wicker et al. 2007). However, SBTEs cannot be classified into the known groups. Therefore, SBTEs should be classified as a new group of transposable elements.

The Bm1 element itself is transcribed by RNA polymerase III. Heat shock and exposure to several toxic compounds increase the level of Bm1 RNA (Kimura et al. 1999, 2001). However, in the permanent cell lines, Bm-5 and BmN, approximately 80% of transcripts containing Bm1 sequences are produced by RNA polymerase II. Moreover, the sizes of most transcripts containing Bm1 sequences are larger (1700–3500 bp) than the Bm1 consensus sequence (429 bp; Gao and Herrera 1996). This raises the possibility that some Bm1 elements are transcribed as part of larger transcripts containing mRNA by way of ‘read-through’ and may be involved in the post-transcriptional regulation of gene expression (Gao and Herrera 1996). However, the functions of these longer transcripts containing Bm1 sequences are not clear. We believe that there is a possibility that these longer transcripts containing Bm1 sequences may be derived from SBTEs.

LTR-type non-autonomous retrotransposable elements, despite their apparent lack of functional ORFs, can propagate using the machinery of other elements (Kalendear et al. 2004; Antonius-Klemola et al. 2006; Sabot et al. 2006). Therefore, we initially thought that Bm1 sequences of Rinne-L, Rinne-S, and Yaocho were carried on each LTR-type retrotransposable element merely as a package. However, our results strongly indicate that LTR-type SBTEs acquired the ability to propagate after these structures containing the Bm1 sequence (figure 1, G, H&I) had been constructed.

The role of Bm1 sequences in SBTEs for transcription or transposition is unknown. The length of Bm1 sequences range from short (Bm1modoki) to long (Ins11). It is likely
that the Bm1 sequence has a specific role for transcription or transposition because the sequences of the flanking regions of Bm1 sequences in each SBTE are not homologous to each other.

The generation processes of SBTEs in the genome of *B. mori* are unknown. One possibility is that the insertions of Bm1 sequences are responsible for the generation of SBTEs. However, there are no nucleotide sequences of ancestral states (not containing the Bm1 sequence) of SBTEs in the WGS data of the p50 strain. We need to consider whether the selection takes place during maintenance of the p50 strain. Another possibility is that SBTEs were generated through an unequal crossing over or intrachromosomal recombination between Bm1 elements (Fujii et al. 2008). To clarify these points, investigations of DNA of numerous silkworm strains maintained worldwide and of the wild silkworm *B. mandarina* are required.

**Non-autonomous transposable elements except SBTEs**

We identified non-autonomous transposable elements other than SBTEs. Chikuri, Minichikuri and Kusanagi contain poly(A) or (TGA)3 tails, respectively. Therefore, it is believed that the insertion mechanism of these three elements is the same as that of Bm1. It seems likely that Chikuri and Minichikuri contribute to the elongation of poly(A) stretches in the genome because these two elements are preferentially inserted into poly(A) sequences. In addition, we investigated whether there is a transposable element containing Chikuri, Minichikuri, or Kusanagi similar to the Bm1 sequence of SBTEs. However, we could not find transposable elements containing Chikuri, Minichikuri, or Kusanagi in the WGS and Build2 data.

Nagisa, Sayama, Shirigonomi and Amanoya do not contain a 3′ tail. Moreover, 5′-truncated types of these four elements could not be found in the genome. Therefore, it is believed that the insertion mechanism of these four elements is different from that of Bm1. Based on the copy number of these four elements, it is believed that the insertion mechanism of these elements may retrotranspose. These elements are not deletion derivatives of other transposable elements and may be classified as a new group of transposable elements.

Kohichi and Supermite-BMC1 have TIR, respectively. Although the role of the endonuclease sequence of BMC1 in Supermite-BMC1 is not known, it is believed that these two elements transpose similar to MITE.

**Functional hybrid transposable elements**

One of the characteristics of the *Bombbyx* genome is an abundance of transposable elements (Mita et al. 2004). In particular, numerous transposable elements have accumulated on the W chromosome (Abe et al. 2000, 2002, 2005a, 2008, 2010). However, numerous are nucleotide sequences that do not show the characteristics of typical transposable elements (Abe et al. 2005a). We considered the long non-coding sequences (e.g., ChoBm1 and Neet) to be junk DNA. We initially thought that these long non-coding sequences had been generated by complicated insertions of numerous short transposable elements. Surprisingly, these non-coding sequences do not have the nested structure of short transposable elements but instead have the structure of long non-autonomous transposable elements. Moreover, several transposable elements contain the Bombyx SINE Bm1 sequences as parts of the element. The findings of SBTEs provide new insights not only into the role of SINE in the host genome but also into the generation of new transposable elements. In particular, the LTR-type SBTEs would be considered degraded LTR retrotransposons if classified using the conventional classification system. Hybrid transposable elements are commonly seen in the nucleotide sequences of Bombyx W chromosome (Abe et al. 2005a) as well as in the genomic sequences of plants (SanMiguel et al. 1988; Shirasu et al. 2000). However, it seems likely that the functional hybrid transposable element is extremely small (Tanskanen et al. 2007). Almost all hybrid transposable elements seem to have arisen from nested transposable element integration or by intrachromosomal recombination (SanMiguel et al. 1988; Shirasu et al. 2000). However, Rinne-L, Rinne-S, and Supermite-BMC1 are believed to be functional hybrid transposable elements because these elements contain parts of other non-LTR retrotransposons (TREST or BMC1) as components and are dispersed in the genome maintaining these structures.

Thus, we identified several novel non-autonomous transposable elements in the *Bombbyx* genome. We have not investigated their presence in other insect genomes. Recently in the codling moth *Cydia pomonella*, the W chromosomal DNAs have been isolated and analysed (Fuková et al. 2007; Marec et al. 2010). The results obtained in this study will be useful not only for the next step of the *Bombbyx* genome analysis, but also for the genome analyses of other organisms.

Acknowledgments

This work was supported by grants from MEXT (no. 17018007 to T. S. and no. 21580063 to H. A.), MAFF-NIAS (Agrigenome Research Programme), NBRP (National BioResource Project), the programme for PROBRRAIN (Promotion of Basic Research Activities for Innovated Biosciences), and JST (Professional Programme for Agricultural Bioinformatics), Japan.

References


Abe H., Kanehara M., Terada T., Ohbayashi F., Shimada T., Kawai S. et al. 1998a Identification of novel random amplified polymorphic DNAs (RAPDs) on the W chromosome of the domest-


Silkworm sex chromosome transposable elements


Received 5 August 2009, in revised form 4 May 2010; accepted 24 June 2010
Published in the Web: 6 September 2010