Available online at www.sciencedirect.com
SCIENCE DIRECT®
Insect Biochemistry and Molecular Biology 35 (2005) 921–929
Insect Biochemistry and Molecular Biology
www.elsevier.com/locate/ibmb

Survey of long terminal repeat retrotransposons of domesticated silkworm (*Bombyx mori*)

Xu Jin-Shan[a], Xia Qing-You[a], Li Jun[b], Pan Guo-Qing[a], Zhou Ze-Yang[a,c,*]

[a]*The Key Sericultural Laboratory of Agricultural Ministry of China, Southwest Agricultural University, Chongqing 400716, China*
[b]*Beijing Institute of Genomics, Chinese Academy of Sciences, Peking 101300, China*
[c]*Chongqing Normal University, 400047, China*

Received 1 November 2004; received in revised form 21 March 2005; accepted 29 March 2005

Abstract

Long terminal retrotransposons are major components of eukaryotic transposable elements. We have surveyed the long terminal repeats (LTR) retrotransposons of domesticated silkworm (*Bombyx mori*) by mining the data produced by *Bombyx mori* Genome Sequencing Project. At least 29 separate families of LTR retrotransposons are identified in this survey, comprising of 11.8% of the complete sequence. Families of domesticated silkworm LTR retrotransposons can be mainly classified into three groups: *gypsy*-like, *copia*-like, *Pao-Bel*. Fourteen families identified consist of *gypsy*-like elements, four families consist of *copia*-like elements and seven families consist of *Pao-Bel* elements. In addition to the three groups of LTR retrotransposons, two families of unusual non-coding elements are identified in the genome of this species. Further phylogenetic analysis of RT domain indicates that the elements of *B.mori* show high diversity and can form different clades in each group. An analysis of sequence variation from different families reveals distinct patterns of variation for the elements belonging to three groups. The analysis of the domesticated silkworm LTR retrotransposons should assist in our understanding of the roles of retroelement in lepidopteron insect genome evolution.

Keywords: Domesticated silkworm; LTR retrotransposons; Sequence variation; Genome

1. Introduction

Retrotransposons are mobile elements that replicate via an RNA intermediate (Boeke and Stoye, 1997). They are the most widespread and enriched class of eukaryotic transposable elements. More than 40% of the genome in human and more than 50% of the genome in maize are comprised of retrotransposons (Smit, 1999; SanMiguel et al., 1996). Retrotransposons can be classified into two categories usually: LTR retrotransposons, and non-LTR retrotransposons. LTR retrotransposons and retroviruses are flanked by long terminal repeat (LTR) in both ends, and they are nearly identical in structure (Xiong and Eickbush, 1990). Non-LTR retrotransposons (also known as LINEs) are those which lack such LTR. LTR retrotransposons containing transcriptional regulatory sites and flanking the two LTRs make up a very large class of transposable elements. For instance, about 8% of the human genome and 10% of the mouse genome are known to be composed of LTR retrotransposons (Lander et al., 2001; Waterston et al., 2002). In previous reports, it is indicated that the biological significance of retrotransposons ranges from their contributions to genes and their postulated roles in genome evolution in both animals and plants (Kumar and Bennetzen, 1999).

In domesticated silkworm, only several LTR retrotransposons elements are identified in previous studies,

*Corresponding author. The Key Sericultural Laboratory of Agricultural Ministry of China, Southwest Agricultural University, Chongqing 400716, China.
Tel.: +86 23 68251088; fax: +86 23 68251128.
E-mail addresses: xjsh@swau.cq.cn (X. Jin-Shan), zyzhou@swau.cq.cn (Z. Ze-Yang).

0965-1748/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ibmb.2005.03.014

and it is suggested that long-terminal-repeat (LTR) retrotransposons could be divided into three major groups (or families), namely the *copia*-like, *gypsy*-like, and *Pao-Bel* groups (Abe et al., 2001); and several LTR retrotransposons elements were also identified in W chromosome (Abe et al., 2000) .The recent release (December, 2004) of the domesticated silkworm genomic sequence with a haploid content of 427.85 Mb (Qingyou et al,. 2004) offers a good opportunity to study the structures and distribution of retrotransposons in domesticated silkworm genome, and undoubtedly, it will also give help to investigate the contribution of LTR retrotransposons to genome evolution in lepidopteran insects.

In this study, we present the result of recent survey of domesticated silkworm genomic data for the presence of LTR retrotransposons by using a search program LTR_STRUC (LTR retrotransposons structure program) and BLAST program as the initial data-mining tool (McCarthy and McDonald, 2003; Altschul et al., 1990). Full-length LTR retrotransposons elements were defined as ones with two LTRs and a pair of target site duplications (TSDs). In order to classify all the elements identified in our survey into distinct families, we define an LTR retrotransposons as a family with similarity of RTs at least 90% at the amino acid level (Bowen and McDonald, 2001). In our study, we also found many non-coding elements that lack an RT ORF, thus they are assigned to the same family on the basis of their same structure and high similarity of LTRs, according to the means used in previous study (Eugene et al., 2002). Taking full-length LTR retrotransposons elements as canonical sequence, we estimated the copy numbers and analyze the sequence variation of distinct families.

By now, there are no consistent reports with regard to domesticated silkworm retrotransposons nomenclature. In our study, LTR retrotransposons are designated as appellation *BmRT* (*Bombyx mori* retrotransposons). Distinct families are denoted by number (for example, *BmRT1*, *BmRT2*, *BmRT3*…). To keep the systemic consistency, we have chosen to adopt the *BmRT* nomenclature in this study to refer to a previously named family.

2. Materials and methods

2.1. Characterization of LTR retrotransposons

We have mined the *B.mori* dataset for LTR retrotransposons using the software LTR_STRUC, which characterize the new LTR retrotransposons based on the existence of characteristic retroelement features such as TSRs, canonical dinucleotides, PBS, PPF and so on. After elements were identified primarily, sequence analyses were performed to find the open read frames (ORFs) encoding reverse transcriptase and other retrotransposons proteins to confirm the retroelement presented in our study. In order to characterize additional elements which are not involved in the initial survey with LTR_STRUC, we have used representative sequences from silkworm database of the Genbank as queries to conduct BLAST (Altschul et al., 1990) searches against our domesticated silkworm genomic database. Thus, by this means, our research can offer the general and reasonable survey of LTR retrotransposons diversity in domesticated silkworm. We use the RepeatMasker programs run under the sensitive condition (Smit and Green, 1996). This program makes us identify all the members covering full-length and partial copies in the whole domesticated silkworm genome by using the complete LTR retrotransposons elements as Repbase, including the elements identified in our research and former research. And in our survey, any elements less than the length of full-length copies are defined as partial. All high-scoring pairs (HSPs) with the length longer than 100 nucleotides and identities higher than 90% were used to define different copies in the same family.

2.2. Multiple sequence alignments and phylogenetic analyses

The RT domains of the distinct *BmRT* elements were identified according to previously described criteria (Xiong and Eickbush, 1988; Xiong and Eickbush, 1990), and these RT domains were aligned with previously reported RT sequence (Table 2), Clustalw (Thompson et al., 1997) software was used to generate amino acid sequence alignments. Neighbor-joining approach was used for phylogenetic tree and bootstrap analysis through the phylogenetic tree menu of Clustalw, and 100 data replicates was generated to analyze bootstrap values. All trees generated were visualized with TreeView and NJPLOT (Page, 1996; Perrière and Gouy, 1996).

2.3. Sequence variation analyses

Transposable elements are considered as a source of spontaneous variation and mutation due to transposition (Bennetzen, 2000). In an initial effort to investigate the patterns of sequence variation of *B.mori* LTR retrotransposon, we use the single nucleotide polymorphism (SNP) to investigate the intra-family variation of different copies of LTR element. Assuming that canonical elements represent active full-length copies, any elements less than the length of full-length copies are defined as partial. Depending on this criterion, cross_match (Phil, 1994) software was used to identify the numbers of SNP by doing the complete pairwise