NUCLEOSOMAL ORGANIZATION OF DROSOPHILA RETROTRANSPOSON INSERTION SITES

* Katokhin A.V., Furman D.P., Levitsky V.G., Katokhina L.V.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia, e-mail: katokhin@bionet.nsc.ru
*Corresponding author

Key words: D. melanogaster, retrotransposon, insertion site, nucleosomal potential, nucleosome context

Resume

Motivation: It is known that nucleosomal organization of chromatin plays an important role in regulation of gene expression. Changes in nucleosome positioning and the consequent changes in expression of genes in the region of insertions are a possible aftermath upon insertion of retrotransposons (REs) into the host genome. This aspect of the interactions between the genome and retrotransposons yet remains vague. The goal of this work is to clarify, first, whether the nucleosome potential profile is specific of the genomic regions housing potential insertion sites for various retrotransposons and, second, whether the nucleosome potential profiles of LTRs of retrotransposons belonging to different families and the corresponding genomic regions encompassing the insertion sites are different.

Results: Nucleosome potential profiles of insertion sites (genomic regions where insertions were detected) for retrotransposons from eleven families were for the first time studied using computer tools and compared with the corresponding profiles of potential insertion sites (“native” sequences obtained by in silico reconstruction). It was demonstrated that (1) potential insertion sites of several REs displayed specific nucleosome profiles correlating with nucleotide compositions of both the target sites themselves and the inverted terminal repeats (ITRs), forming the 5’ and 3’ ends of each retrotransposon LTR, and (2) the profiles of nucleosome potentials of retrotransposon LTRs from different RE families differed from one another as well as from the profiles of the genomic regions encompassing these inserts.

Introduction

Mobile elements (transposons) constitute a considerable part of the eukaryotic genomes. The mobile elements in general and retrotransposons (REs) in particular represent a mighty source of genotypic variation due to their influence on the gene structure/function organization in the regions of insertions.

As a rule, the main attention of the corresponding research is focused on the effects of RE insertions on the gene structure–function organization stemming from the changes in primary DNA sequence. However, it is known that expression of genes depends as well on chromatin packaging (Bonifer, 1999). Thus, it is natural to assume that REs are capable of changing function of the genes localized to the regions of insertions not only via the interactions at the DNA level, but also in an indirect manner through modulating the chromatin structure, in particular, its nucleosomal organization. In turn, this suggests that the nucleosomal organization of a DNA fragment wherein a certain retrotransposon has inserted might also be of importance to the integration itself.

Availability of Drosophila complete genomic sequence in combination with adequate computer methods for detecting and analyzing contextual and conformational properties of DNA sequences allowed these questions to be studied theoretically. The goal of this work was (1) to find out whether the nucleosome potential (NP) profiles of genomic regions containing potential sites for insertion of retrotransposons from various families are specific and (2) to detect the differences between the NP profiles of various retrotransposon LTRs and those of genomic regions encompassing the inserts.

Materials and Methods

The reconstructed euchromatic sequence of D. melanogaster genome (Adams et al., 2000) and the RE sequences with the copy numbers exceeding seven retrieved from FlyBase (http://flybase.harvard.edu:7081/transposons/lf/melanogaster-transposon.html) were used for the analysis.

The REs were localized in the genomic sequence by the program BLASTn (http://www.ncbi.nlm.nih.gov/BLAST/ and http://www.fruitfly.org/blast/index.html) using the default parameters.

The NP profiles were studied using sets of sequences with a length of 1200 bp (±600 bp relative to the site of RE insertion) divided into two following groups: (1) “5’flanks + LTRs”, comprising genomic sequences adjacent to left LTRs together with the LTR sequences, and (2) “5’flanks + 3’flanks”, comprising genomic fragments adjacent to left LTRs attached to the sequences adjacent to right LTRs. The total number of analyzed sequences for REs belonging to 11 families (Table) amounted to 328. The method RECON (Levitsky et al., 2001) was used to calculate NP.
Table. Characteristics of D. melanogaster retrotransposons analyzed.

<table>
<thead>
<tr>
<th>RE family</th>
<th>Target site</th>
<th>ITR</th>
<th>LTR (bp)</th>
<th>Copy number</th>
</tr>
</thead>
<tbody>
<tr>
<td>297</td>
<td>ATAT</td>
<td>AG/CT</td>
<td>415</td>
<td>20</td>
</tr>
<tr>
<td>17.6</td>
<td>ATAT</td>
<td>AG/TT</td>
<td>512</td>
<td>7</td>
</tr>
<tr>
<td>yoyo</td>
<td>T(At)(Ta)(A)</td>
<td>AG/CT</td>
<td>519</td>
<td>8</td>
</tr>
<tr>
<td>HMSBeagle</td>
<td>AG(A)(Tc)(A)</td>
<td>AG/CT</td>
<td>266</td>
<td>7</td>
</tr>
<tr>
<td>mdg3</td>
<td>(Gc)(Tc)(A)</td>
<td>TG/CA</td>
<td>492</td>
<td>8</td>
</tr>
<tr>
<td>Dm412 (mdg2)</td>
<td>(Ag)(Tg)(At)</td>
<td>TG/AA</td>
<td>267</td>
<td>7</td>
</tr>
<tr>
<td>copia</td>
<td>(Ag)(Ta)(Ta)</td>
<td>TG/CA</td>
<td>276</td>
<td>14</td>
</tr>
<tr>
<td>blood</td>
<td>(Gc)(Ag)(Gc)</td>
<td>TG/CA</td>
<td>400</td>
<td>13</td>
</tr>
<tr>
<td>roo (B104)</td>
<td>(Agc)(Tcg)(Tc)(Agg)(Cg)</td>
<td>TG/CA</td>
<td>429</td>
<td>36</td>
</tr>
<tr>
<td>tirant</td>
<td>CGCG</td>
<td>AG/CT</td>
<td>417</td>
<td>10</td>
</tr>
</tbody>
</table>

Results and Discussion

Comparative analysis of the data obtained demonstrated a diversity of NP profiles of the sequences from both sets. Most typical profiles are shown in Fig.

Fig. NP profiles of insertion sites of (A) 17.6, (B) mdg3, (C) Dm412, (D) roo, and (E) tirant retrotransposons: the ordinate, value of NP; the abscissa, position (bp); the insertions correspond to position 600; gray line, NP profiles of group (1) “5’flanks + LTRs”; black line, NP profiles of group (1) “5’flanks + 3’flanks”; arrows above NP profiles indicate LTR lengths; and right-angle arrow marks the transcription start for mdg3 (B).
1) Properties of “5’flanks + 3’flanks” sequences:

The sequences of this type free of 17.6 and 297 inserts (17.6-less and 297-less) displayed a decrease in the value of NP in the region of ±100 bp relative to the site center (Fig. A). The analogous tirant sequences (Fig. E) manifested this property to a lesser degree. The sequences free of the rest REs did not display this pattern.

The elements 17.6 and 297 have identical IT Rs and target sites (Table), suggesting that this underlies the observed similarity in their nucleosomal potential profiles.

The tirant ITR displays the same composition but a different target site (CGCG instead of ATAT). However, both sites are purine–pyrimidine motifs; hence, they are similar in certain conformational features (Fitzgerald, Anderson, 1999). Consequently, the specific features of NP profiles stemming from the presence of purine–pyrimidine tracts are likely to be important for integration of the REs with ITRs of AG/TC composition into the genome.

2) Properties of “5’flanks + LTRs” sequences:

The “5’flanks + LTRs” sequences with 17.6 (Fig. A), 297, and copia (data not shown) REs exhibit elevated NP over their LTR fragments compared with the corresponding sequences of “5’flanks + 3’flanks” type, that is, with the encompassing genomic context. This inference is experimentally confirmed for copia: it was demonstrated that a high nucleosome density of a copia sequence is comparable with the density typical of heterochromatin (Sun et al., 2001).

On the contrary, LTRs of all the rest REs - mdg3, Dm412, roo, tirant (Fig. B–E), yoyo, mdg1, blood, and HMSBeagle (data not shown)—display decreased values of the NP within the inner LTR regions. The NP profile observed reflects the presence of two nucleosomes at both LTR ends and an intermediate region of “open” chromatin. As a rule, such nucleosome-free regions correspond to gene regulatory regions, in particular, promoters (Bonifer, 1999; Levitsky et al., 2001). Promoter of mdg3 is actually located precisely in this region (Mazo et al., 1986). The LTRs of the retroviruses MMTV and HIV-1 appeared to exhibit similar architecture (Bonifer, 1999; Marzio, Giacca, 1999).

References