Retrotransposon gypsy and Its Role in Genetic Instability of a Mutator Strain of Drosophila melanogaster

N. V. Lyubomirskaya¹, A. I. Kim², and Yu. V. Ilyin¹

¹ Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 119991 Russia; fax: (095) 135-14-05; e-mail: ilyin@genome.eimb.relarn.ru
² Department of Genetics and Breeding, Biological faculty, Moscow State University, Moscow, 119899 Russia; fax: (095) 939-35-12; e-mail: kim@akim.home.bio.msu.ru

Received October 4, 2002

Abstract—This article summarizes the results of a ten-year study of genetic instability of a mutator strain of Drosophila melanogaster caused by transposition of the gypsy retrotransposon. The results of other authors working with an analogous system are analyzed. Possible mechanisms are suggested for the interaction of gypsy with the cell gene flamenco that participates in transposition control of this mobile element.

INTRODUCTION

Since the works by Britten and Cohen that have laid the foundations of studying the eukaryotic genome by means of DNA reassociation kinetics, when analysis of individual genes seemed a problem of distant future, DNA redundancy has been an intriguing issue attracting the attention of researchers. Even before the era of gene engineering, it was shown that over 70% of the genomic DNA is represented by repeats including highly and mid-repetitive sequences. After the advent of DNA reassociation kinetics, when analysis of individual regions of genomic DNA and the discovery of mobile genetic elements, these elements were shown to constitute a significant part of the middle repeats.

The existence of mobile genetic elements has been predicted by Barbara McClintock as early as in 1951 [1] but her discovery was recognized and eventually awarded the Nobel Prize only in 1983. This was possible due to the discovery of mobile genetic elements at the molecular level and the isolation of Drosophila retrotransposons that were made simultaneously by the teams of G.P. Georgiev [2] and D. Hogness [3]. The following decade has witnessed a boom in mobile genetic element studies resulting in significant discoveries, many of which were made by Georgiev and his associates and students.

Today the main lines of retrotransposon investigation are their interaction with other cell genome components and their role in microevolution. Of special interest are systems with weak genetic control leading to genetic instability. These systems are a good model for studying mechanisms of interaction of eukaryotic genome components. One of these systems is the Mutator Strain (MS) derived from the Stable Strain (SS), which is among oldest laboratory strains of D. melanogaster from the Moscow State University collection [4].

The Mutator Strain exhibits high mutation rate in germline and somatic cells (up to 10⁻³–10⁻⁴ and 10⁻²–10⁻³, respectively), frequent appearance of instable mutations of different types, premeiotic mutation [4], high rate of chromosome aberrations in larval neuroblasts [5], and altered behavior of adult flies [6]. The present paper summarizes the results of the studies of the last decade that analyze genetic instability of MS at the molecular level. This system is of obvious interest because it exhibits selective amplification and transposition of retrotransposon gypsy at the background of the unchanged localization of other mobile elements [4]. We would like to remind that gypsy was first discovered and described by the team of Georgiev [7], to whose anniversary is dedicated this issue of the Russian Journal of Genetics. An apparent advantage of this system is the existence of two laboratory strains, one of which was derived from the other acquiring genetic instability, which permits their comparative analysis [8].

THE NATURE OF GENETIC INSTABILITY OF THE MUTATOR STRAIN

Although genetic instability can have various causes, two explanations seem most plausible in the present case: alterations in the structure of the mobile genetic elements itself leading to its selective mobilization and alterations in the regulatory systems of the strain affecting gypsy transposition. The first of these two possibilities was the subject of our early studies.

Using molecular analysis, two gypsy subfamilies have been identified [9, 10]. These subfamilies exhibit marked structural differences detected even at the restriction level (Fig. 1). Note that the presence (absence) of an internal HindIII sites is always associated with the presence (absence) of XbaI sites in the long terminal repeats (LTRs). This means that the observed differences reflect the existence of two sub-
families rather than simple structural polymorphism at individual sites. This divergence of gypsy into two subfamilies may reflect evolutionary processes favoring their further division and formation of new families. In Drosophila, families of retrotransposons 17.6 and 297 [11], as well as MDG1 and 412 [12] are known, which are likely to descend from one ancestor since they have pronounced homologies.

A question arises why gypsy copies from the first subfamily were found in the unstable strain while lacking in the original, stable strain. This is explained by the history of the mutator strain. The study of this strain initially was not aimed at investigating genetic instability [13]. Because of this, the creation of this strain involved, after treatment with a mutagen, crosses with other laboratory strains regardless of mobile genetic elements contained in them. It is possible that at this stage, gypsy copies from the “active” subfamily were introduced into the genome of the strain in question. Genetic instability appeared in the strain at a later stage.

The presence of two gypsy variants was also noted earlier. One of the variants prevailed in the strains that did not exhibit gypsy transposition whereas the other predominated in Drosophila cell culture and in the strains showing amplification and mobilization of gypsy [7]. Moreover, all gypsy copies cloned from the mutant loci [14–17] or from cell culture [7] were represented by only one variant corresponding to that prevailing in the cell culture. These results suggested that two gypsy variants differed not only in structure but also in retrotransposition activity. However, the complete identity of the left and right LTRs in the p7K copy cloned from SS as well as the absence of alterations in the inverted repeats at the LTR ends significant for transposition indicate that the members of the second subfamily are in principle capable of transposition [9, 18]. The members of this subfamily do not differ from transpositionally active p6K copies by the structure of regions important for transcription and by their transcriptional activity upon transient expression in cell culture [9]. This also testifies to the fact that they are more than just defective copies. Nevertheless, we did not succeed in detecting gypsy transcription in SS. This may be accounted for by specific regulation of gypsy transcription in the stable strain, the extremely small number of its copies (only two euchromatic sites) and/or the impact of the chromosomal background on their expression. The third explanation seems most plausible.

Apparently, the differences in transposition activity have post-transcriptional determination. They may be caused by differences in the protein products of the mobile element itself. To clarify this issue, a detailed analysis of the complete primary structure of the both copies was conducted. This analysis showed that the structural differences between the copies constitute only 0.2% of the coding region. All the differences were represented by nucleotide substitutions, mostly at the third codon position. The total number of amino-acid substitutions was 14 (Fig. 2) [10, 18].

To determine, which of the detected structural differences cause the functional differences between the two variants, we developed a model system that allowed us to trace the appearance of new gypsy copies formed via reverse transcription. It was shown that the two gypsy subfamilies actually differ by their activity [19]. To reveal concrete structural differences between the gypsy variants accounting for their functional difference, we carried out complete sequencing of the “inactive” element and created hybrid constructs that contained the “active” and “inactive” gypsy copies in different combinations [10, 20].

Two amino-acid substitutions were found: Thr-Asp (nucleotide substitution 2928) and Lys-Arg (nucleotide substitutions 3907, 3908). One or both of them directly affect retrotransposition efficiency [10, 20]. These amino-acid substitutions are located in the second open reading frame of gypsy. One of them is at the boundary of domains homologous to retroviral protease and revetase. Thus, this substitution may influence processing of enzyme products of the retrotransposon. The other of the amino-acid substitutions is located in the second of seven revetase domains and, consequently, may directly affect the enzyme activity. To clarify this issue, the revertase activity of the both gypsy variants must be analyzed in more detail.

**Fig. 1.** Restriction map of gypsy: 1, retrotransposon LTR; 2, leader sequence; 3, coding sequence. Restriction sites: RI, EcoRI; K, KpnI; G, BglII; Xb, Xbal; Ml, MluI; H, HindIII; P, PstI; Pv, PvuII; X, XhoI; C, ClaI. Restriction sites absent in the “inactive” copy are marked with asterisks; the numerals show the arrangement of restriction sites in the gypsy nucleotide sequence (correspond to the first nucleotide in the recognized sequence).