Distribution and sequence homogeneity of an abundant satellite DNA in the beetle, \textit{Tenebrio molitor}

Claytus A. Davis* and Gerard R. Wyatt

Department of Biology, Queen's University, Kingston, Ontario K7L 3N6, Canada

Received March 17, 1989; Revised and Accepted June 13, 1989

ABSTRACT

The mealworm beetle, \textit{Tenebrio molitor}, contains an unusually abundant and homogeneous satellite DNA which constitutes up to 60\% of its genome. The satellite DNA is shown to be present in all of the chromosomes by \textit{in situ} hybridization. 18 dimers of the repeat unit were cloned and sequenced. The consensus sequence is 142 nt long and lacks any internal repeat structure. Monomers of the sequence are very similar, showing on average a 2\% divergence from the calculated consensus. Variant nucleotides are scattered randomly throughout the sequence although some variants are more common than others. Neighboring repeat units are no more alike than randomly chosen ones. The results suggest that some mechanism, perhaps gene conversion, is acting to maintain the homogeneity of the satellite DNA despite its abundance and distribution on all of the chromosomes.

INTRODUCTION

The genomes of almost all metazoan species contain satellite DNAs, sequences repeated thousands of times in tandem arrays. A considerable body of data concerning their nature and sequence in many species has accumulated (1,2,3). In general, the sequences are not transcribed, never encode proteins and are almost invariably associated with heterochromatin. Despite extensive analyses no function has yet been demonstrated for satellite DNAs, although many have been suggested (4,5).

Satellite sequences are interesting from an evolutionary point of view. Different satellites vary widely in sequence and repeat length, but despite great sequence divergence between closely related species, the repeats within a species are usually quite similar, a phenomenon that has been attributed to concerted evolution. The wide sequence variation amongst even related species and the lack of transcription suggest that any function they may have does not depend on retention of a particular sequence. Study of satellite DNAs may, therefore, be a way to examine mechanisms of genome dynamics uninfluenced by sequence-specific selection pressure. For this reason, extensive analysis of satellite DNAs will be useful.

In this study we have analyzed the distribution and sequence variation of an unusually abundant and homogeneous satellite DNA in the beetle, \textit{Tenebrio molitor}. While this manuscript was in preparation a description of the \textit{T. molitor} satellite DNA was published (6).

MATERIALS AND METHODS

\textit{Insects}: Mealworm beetles, \textit{Tenebrio molitor} L., were obtained from a local dealer and maintained on bran meal.

\textit{DNA isolation}: DNA was isolated by a detergent lysis method, similar to those described
DNA restriction fragments were isolated from agarose gels by electrophoresis onto a DEAE cellulose membrane (NA-45, Schleicher and Schuell) and eluted according to the supplier's instructions.

**Cloning:** *T. molitor* genomic DNA was partially digested with *EcoRI* and monomers and tetramers of the repeat unit were cloned into the *EcoRI* site of pUC8. Dimers were also cloned into the single stranded sequencing vector M13mp8.

**In situ hybridization to metaphase chromosome spreads:** Day 5 male pupae were injected with 3 μl of 1% colchicine. After 1.5 hours, accessory glands were isolated (8), and then treated with Dispase (10 mg/ml, 10 min at room temp.) to produce a cell suspension. Metaphase spreads were made from these cells after swelling (75 mM KCl, 10 min) and fixing with methanol:acetic acid (3:1). The slides were washed in 2× SSC and then dehydrated through an ethanol series. The chromosomes were denatured in 70% formamide-2× SSC, pH 7.0 for 2 min at 70°C. Hybridization solution (50 μl of 25% formamide, 5% dextran sulphate, 2× SSC, 0.5 mg/ml denatured sheared herring sperm DNA and 0.2 ng/ml probe) was then applied and the slides were incubated for 12 hours at 37°C. The cover slips were removed and the slides were washed 3 times for 3 min each in 50% formamide-2× SSC at 39°C and 5 times in 2× SSC. They were then dehydrated through an ethanol series, dipped in a 50% solution of Kodak NTB2 emulsion at 45°C, dried at room temp, for a few hours and exposed at 4°C. Metaphase chromosome preparations were stained with Wright's stain.

**DNA sequencing:** Single stranded M13 clones were sequenced by the dideoxy procedure (9) as modified for 35S labelling (10). After electrophoresis the gels were dried and exposed for 4—7 days at −70°C. To minimize sequence reading errors, an automatic computer sequence entry program was used. The resulting sequence data were searched for restriction enzyme sites and secondary structures with the Queen and Korn sequence analysis program (11).

**RESULTS**

An abundant satellite component in the DNA of *Tenebrio molitor*

Digestion with the restriction endonuclease *EcoRI* converted a substantial portion of *Tenebrio molitor* genomic DNA into fragments of approximately 140 bp and multiples thereof. The relative abundance of these fragments, separated by electrophoresis in 2% agarose gel and measured by densitometric scanning of photographic negatives, indicated that they represented approximately 60% of the genome (data not shown). In a similar analysis a value of 44 to 56% was obtained (6). The diploid genome of *Tenebrio molitor*, estimated as 1.9 pg of DNA (8), therefore contains approximately 7.5 million copies of the repeat unit. In a Southern analysis of genomic DNA digested with *EcoRI* and probed with cloned repeat elements (final washing stringency—0.1× SSC, room temp.) very little hybridization was observed outside of the repeat ladder (data not shown). This result indicates that there is very little DNA that is more than 65% identical to the repeat sequence but does not produce a repeat-sized band after *EcoRI* digestion.

Partial digestion with *EcoRI* produced a long ladder of bands in gel electrophoresis, suggesting that the repeated units were arranged in tandem in the genome and therefore represented a satellite DNA (data not shown). By equilibrium centrifugation in CsCl in the presence of the dye Hoechst 33258, the satellite DNA was resolved as a sharp band with greater buoyant density than the remaining genomic DNA (not shown). Northern analysis of total larval RNA showed no evidence for transcription of the satellite (not shown).
Figure 1. Distribution of the repeated sequence on the chromosomes of T. molitor. Metaphase spreads of chromosomes from accessory glands of 5 day male pupae were prepared and hybridized to a $^3$H-labelled cloned tetramer of the repeat. The specific activity of the probe was $3.5 \times 10^7$ dpm/µg and the exposure time was 27 hours at 4°C. a,b—two different metaphase spreads.

Distribution of the Satellite in the Genome

A tritium-labelled cloned tetramer of the repeat was hybridized in situ to metaphase spreads of T. molitor chromosomes (Fig. 1). Over 30 metaphase spreads were examined. All twenty chromosomes including the minute Y chromosome hybridized to varying degrees. Control slides with metaphase spreads of T. molitor probed with a human-specific probe, and human metaphase spreads probed with the Tenebrio repeat, showed no hybridization. Treatment with RNAse-free DNase completely abolished labelling. Because of the small sizes of the T. molitor chromosomes and degradation during the hybridization procedure we could not identify the chromosomes or estimate differences in their contents of the satellite DNA.

However, it has been reported that each of the 20 chromosomes possesses a single large block of centromeric heterochromatin of approximately equal size which collectively account for 38% of their length (12), and it is most likely that these contain the satellite DNA.

Estimation of sequence homogeneity by restriction enzyme analysis

The frequency of the six nucleotide sequence comprising the EcoRI restriction site was estimated by determining by laser densitometry the amount of DNA in the monomer-

Figure 2. Size variation in the repeat. 200 ng of T. molitor genomic DNA was digested with EcoRI and end-labelled with $[^{35}S]dATP$ and Klenow enzyme. Serial dilutions were then electrophoresed through a polyacrylamide gel along with a set of sequencing reactions to mark single nucleotide spacing. Approximate amounts of genomic DNA loaded (in pg) are indicated.
The consensus sequence of the T. molitor repeat unit as derived from the compilation of data from 16 repeat dimers is listed starting with the first nucleotide of the EcoRI site. Two regions of dyad symmetry are indicated with arrows.

Figure 3. Consensus sequence of the repeat unit. The consensus sequence of the T. molitor repeat unit as derived from the compilation of data from 16 repeat dimers is listed starting with the first nucleotide of the EcoRI site. Two regions of dyad symmetry are indicated with arrows.

dimer- and trimer-sized bands observed after EcoRI digestion and agarose gel electrophoresis (data not shown). Assuming random alteration of EcoRI sites with a probability of \((1 - p)\), the probability that any particular nucleotide of the repeat lies within an n-mer band is \(n(1 - p)^{n-1}p^2\). In such a random model the intensity ratio of monomer to dimer bands is \((1 - p)/2\) and that of dimer to trimer bands is \(2(1 - p)/3\). The monomer/dimer ratio obtained by scanning complete EcoRI digests after gel electrophoresis was 4.9. The value of \(p\) is then 0.9, and the theoretical dimer/trimer ratio is 6.7, close to the observed value of 8.3. If the variation at the EcoRI site is representative of the entire repeat sequence, then individual units are, on average, 98% identical to the consensus. This estimate is approximate because of possibly incomplete EcoRI digestion and non-linear photographic film response.

To estimate the occurrence of insertion/deletion mutations in the repeat, serial dilutions of genomic DNA cut with EcoRI and then end-labelled were examined by autoradiography after electrophoresis through a 7.5% acrylamide sequencing gel (Fig. 2). The resulting major band of approximately 140 nt is flanked by weaker satellite bands one nucleotide larger and smaller. These bands are broader than those from the sequencing reactions possibly because the two strands may have slightly different mobilities. The ratio of the intensities of the satellite bands to the main band is approximately 0.1, which suggests that the frequency of single nucleotide insertion and deletion events is approximately \(0.1/140 = 7 \times 10^{-4}\) events/position. Overexposure of the gel did not reveal any bands within 20 nt of the monomer that could represent additions or deletions of short sequences in the repeat.

Sequence analysis of the T. molitor satellite

The homogeneity of the satellite DNA was examined in more detail by sequence analysis. Total genomic DNA was partially digested with EcoRI and 18 dimer sized fragments were cloned and sequenced. Of these, three deviated substantially from the later established consensus sequence (Fig. 3) and were rejected as probable cloning artifacts.

In total, 4.4 kb of the repeat DNA, representing 31 monomers, was used in the analysis. Once the sequences were read and the consensus determined, all variant sites were checked again. An unambiguous consensus sequence (Fig. 3) was derived from the complete data. It is 142 nt long, which agrees closely with the estimate based on the electrophoretic mobility of restriction fragments. The sequence was analyzed using the Queen and Korn computer
Figure 4. Variant positions in the repeat unit. The variant nucleotides in 16 sequenced dimers of the repeated unit are shown. The second half of each dimer (b) is listed directly under the first half (a) starting at the internal EcoRI site. The variant nucleotide sites are numbered at the top, positions showing no variation being omitted. The symbols ‘<’ and ‘>’ mark the first and last nucleotides sequenced in some of the clones. Ambiguous sequence is marked with a ‘?’. The vertical gap between position 69 and 70 is the site of an insertion into the sequence of clone T21.

Program (11). There was no evidence of any substructure in the repeat unit as determined by dot-matrix comparison, distribution of trinucleotides or searches for internal repeats, except for two regions of dyad symmetry that could potentially form snapback secondary structures (Fig. 3). To look for patterns of deviation, sequence which conformed to the consensus was suppressed, leaving only the variant nucleotides (Fig. 4). In a total of 4402 nt, there were 89 variant nucleotides. Of these, 88 were substitutions and only one was an addition/deletion variant. The overall sequence variation of 2% was in agreement with the estimate of the EcoRI site abundance. As expected from the size distribution analysis of the monomer, insertion/deletion variants were rare.

Many of the deviations were obviously non-random alterations. For instance, position 11 was approximately 50% G and 50% A; position 63 showed five variants (all of them C) in the 31 monomer sequences; position 79 contained three T variants, position 91 showed three G variants and position 119 showed four A variants. It is possible that others of the variants observed were also non-random but were not recognized as such because of the sample size. The obviously non-random variants accounted for 29 of the total 89 variant positions, and are almost certainly due to the spread of single mutations by whatever processes are responsible for the maintenance of homogeneity in the satellite.

While the frequency of particular variations was often non-random, there was no visible pattern to the distribution of variants. Firstly, all regions of the 142 nt sequence appeared
prone to variation. Secondly, there was no significant difference in the number of variants in different monomers. Thirdly, there was no apparent association between any of the different variants, so that a repeat which possessed any one variant showed no tendency to contain any other particular variant. Finally, the two halves of the dimers sequenced were no more similar to each other than to randomly chosen monomers. In other words, neighboring sequences were no more similar than distant sequences. The sequencing results suggest that the *Tenebrio molitor* satellite is extremely homogeneous, despite its distribution on all of the chromosomes.

**DISCUSSION**

Digestion of DNA from the beetle *Tenebrio molitor* with the restriction enzyme EcoRI released a large portion, approximately 60%, of the genome into 142 bp fragments representing a satellite DNA of exceptional abundance and homogeneity, as recently reported (6). To illuminate the structure and possible evolution of this unusual satellite we examined its chromosomal distribution and sequence variation by three different approaches. First, we measured the variation of the EcoRI site using complete EcoRI digests of genomic DNA; second, we examined the frequency of oligonucleotide insertion/deletion events by acrylamide gel electrophoresis; finally, we sequenced 18 randomly selected dimers. Although our analysis of the *Tenebrio* satellite DNA is restricted to the component which contains EcoRI sites at 142 bp intervals, it is probably representative of the satellite DNA in general since a Southern analysis showed that almost all genomic DNA more than 60% identical to the EcoRI repeat is cut into fragments that are either the same size or multiples of this unit. Our results concerning the abundance and sequence of the repeat are similar to those previously reported (6). The prior estimate of its abundance is 49% compared to our 60% of the genome. The single monomer previously sequenced was 142 bp long and varied from our consensus sequence at positions 39 (T vs. A) and 57 (C vs. T).

Our results show that the *T. molitor* satellite is distributed on all of the chromosomes. Despite this, the monomers are all quite similar to each other, deviating from the consensus by an average of only 2%. The different sequence variants showed no tendency to associate, so it was not possible to group the different monomers on the basis of shared variants. Consecutive monomers had no tendency to share the same variants and were as different from each other as random monomers.

Our data on the *Tenebrio* satellite may be compared with observations on other satellite DNAs. Eight independent multimers of the Sau3A segments of the bovine 1.706 g/cm³ satellite have been sequenced (13). As with the *T. molitor* satellite, there was neither clustering of variants within the repeats nor was there any indication that particular variants were associated with each other. However, the overall level of variation was 12%. The 1.688 g/cm³ sequence of *Drosophila melanogaster* consists of approximately 15,000 copies of a 359 bp monomer, probably restricted to the X chromosome (15). The level of sequence variation found by sequencing 15 consecutive monomers was 3.6% (14). The sequencing of random monomers of the 189 nt EcoRI repeat in *D. gymnobasis* (16) showed that there was a clear association of variants at different positions in the repeat with each other, so that a clone containing one particular variant nucleotide often contained another particular variant. The result was to divide the sequenced clones into groups on the basis of shared variants. Within a group, variation was between 1.5% and 2%, while between groups, it was 12%. Similar analyses of cloned monomers of a rat satellite DNA found on most of the chromosomes (17), calf satellite I (1.715 g/cm³) (18) and a brine shrimp...
satellite (19) also divided these satellites into classes. In the case of the calf satellite, 300 bp of two neighboring repeats that were sequenced were identical (20), suggesting that neighboring sequences were more similar than distant ones. Quite possibly the ability to subdivide the monomers into different classes reflects a hierarchical organization and/or the presence of different variants on different chromosomes. The best characterized satellite DNA, the human alpha satellite, consists of tandem arrays of 170–171 bp monomers which are clearly subdivided into different classes and are organized in hierarchical structures, many of which are specific to one or several chromosomes (21,22).

The *T. molitor* satellite is unusual not for its abundance, homogeneity or occurrence on all of the chromosomes but because of the combination of all three. The most obvious question arising from the sequence variation analysis is, what mechanism has maintained this homogeneity? Of the mechanisms that have been proposed to describe the evolution of satellite DNAs, the unequal crossover model has the most experimental evidence and, in theory, could explain much of the structure of satellite DNAs. Unequal crossover, however, can act only within repeats clustered on the same chromosome and results in neighboring sequences becoming more similar to each other than to distant sequences. In the case of the *T. molitor* satellite, it is difficult to reconcile this mechanism with the observation that random repeat units drawn from different chromosomes are as alike as neighboring ones. The results suggest that some general homogenizing mechanism is at work on these sequences. One such mechanism is gene conversion, which can presumably operate on homologous sequences located anywhere in the genome. Provided that the process is unbiased, it could explain both the homogeneity and the lack of subclasses among the repeat units.

In summary, we have examined the distribution and sequence variation of the exceptionally abundant satellite DNA of the beetle, *T. molitor*. In comparison with satellite DNAs in other species the *T. molitor* satellite is unusually homogeneous. Neighboring repeat sequences were no more alike than random ones and different variants were not associated with each other in individual monomers. These results suggest that some mechanism, perhaps gene conversion, is acting upon the genome of *T. molitor* to maintain homogeneity in the satellite sequences on all of the chromosomes.

**ACKNOWLEDGEMENTS**

We would like to thank Ernest Kroeker for providing the mealworms and advice on them, Clive Morgan and Alessandra Duncan for help with *in situ* hybridization techniques, B.N. White for providing facilities during the initiation of this project, and the Natural Sciences and Engineering Research Council of Canada for a postgraduate scholarship to C.D. and grant support.

*Present address: Mount Sinai Hospital Research Institute, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada*

**REFERENCES**

15. Brutlag, D.L., personal communication in (16)