

## Long Inversely Oriented Subunits Form a Complex Monomer of *Tribolium brevicornis* Satellite DNA

Brankica Mravinac,<sup>1</sup> Đurđica Ugarković,<sup>1</sup> Damjan Franjević,<sup>2</sup> Miroslav Plohl<sup>1</sup>

<sup>1</sup> Ruđer Bošković Institute, Department of Molecular Biology, Bijenička cesta 54, HR-10002 Zagreb, Croatia

<sup>2</sup> University of Zagreb, Faculty of Science, Department of Zoology, Rooseveltov trg 6, HR-10000 Zagreb, Croatia

Received: 29 July 2004 / Accepted: 7 November 2004 [Reviewing Editor: Dr. Willie Swanson]

**Abstract.** Highly abundant satellite DNA named TBREV is detected and characterized in the beetle *Tribolium brevicornis* (Insecta: Coleoptera). An outstanding peculiarity of the TBREV satellite monomer is its complex structure based on the two ~470-bp-long subunits, inversely oriented within a 1061-bp-long monomer sequence. The proposed evolutionary history demonstrates a clear trend toward increased complexity and length of the TBREV satellite monomer. This tendency has been observed on three levels: first as direct and inverted duplications of short sequence motifs, then by inverse duplication of the ~470-bp sequence segment, and, finally, by spread of inversely duplicated elements in a higher-order register and formation of extant monomers. Inversely oriented subunits share a similarity of 82% and have a high capacity to form a thermodynamically stable dyad structure that is, to our knowledge, the longest ever described in any satellite monomer. Analysis of divergences between inversely oriented subunits shows a tendency to a further reduction in similarity between them. Except in its centromeric localization, the TBREV satellite does not show similarity to other known *Tribolium* satellites, either in nucleotide sequence or in monomer length and complexity. However, TBREV shares common features of other *Tribolium* satellites that might be under functional constraints: nonconstant rate of evolution along the monomer sequence, short inverted repeats in the vicinity of an A+T tract, nonrandom distribution of A or T ≥3 tracts, and CENP-B box-like

motifs. Although long inverted subunits might reinforce structural characteristics of the satellite monomer, their nucleotide sequence does not seem to be under constraints in order to preserve the dyad structure.

**Key words:** Inverted repeats — Higher-order register — Heterochromatin — Gene conversion — Concerted evolution — Coleoptera

### Introduction

Tandemly repeated sequences or satellite DNAs represent a major DNA component in heterochromatin, still the most mysterious and least understood genomic compartment. Although once considered nonfunctional, useless sequences accumulated solely because of the low recombination rate in the heterochromatic portion of a chromosome, satellite DNAs have started to attract more attention due to recent results favoring their fundamental importance in the function and evolution of eukaryotic genome. It has been evidenced that long homogeneous domains of satellite repeats, occasionally interrupted with transposable elements, are the principal DNA components involved in centromeric function of animal and plant species, such as human (Schueler et al. 2001), *Drosophila* (Sun et al. 2003), and rice (Cheng et al. 2002). Transcripts of satellite repeats seem to be involved in the formation of densely

packed heterochromatin, mobilizing a number of proteins through the RNAi mechanism (Volpe et al. 2002, Martienssen 2003).

Satellite repeats evolve according to the principles of concerted evolution, in which diverse mechanisms of nonreciprocal transfer induce a high turnover of satellite sequences (Dover 1986). Consequently, profiles of satellite sequences are species specific due to differences in copy number, nucleotide sequence, or composition of satellites (reviewed by Ugarković and Plohl 2002). It is difficult to understand how repetitive sequences unrelated in origin, nucleotide sequence, length, or organizational complexity could share common functions, for example, in the formation of a functional centromere. It has been recently observed that a response to rapid alterations in centromeric satellites is adaptive coevolution of the specific centromeric histone CENH3 (Malik and Henikoff 2001; Talbert et al. 2002).

It is assumed that structural characteristics of a satellite DNA, and not the nucleotide sequence itself, can hold certain functional information. Relatively short, no longer than 30–40 nt, inversely duplicated motifs potentially able to induce formation of stable dyad structures are often detected in evolutionarily unrelated satellite sequences (e.g., Lorite et al. 2004; Castagnone-Sereno et al. 2000). It has been suggested that such structures might be functionally significant in interactions with specific heterochromatic proteins (Bigot et al. 1990; Rojas-Rousse et al. 1993). In addition, functional information of satellite sequences might be conserved in structured helical axis and satellite monomer length (Plohl et al. 1998), in ordered distribution of A + T nucleotides (Barceló et al. 1998), or in sequence motifs, such as the CENP-B box-like motif (Lorite et al. 2004). Analysis of intraspecific sequence variability revealed an alternating pattern of conserved and variable segments along the satellite monomer in *Arabidopsis thaliana* 180-bp centromeric satellite DNA, as well as in human  $\alpha$  satellite DNA, indicating that some sequence segments might be under functional constraints, probably due to the interaction with heterochromatic proteins (Hall et al. 2003). In addition, a similar pattern has been observed in satellite sequences from the beetles *Tribolium anaphe* and *T. destructor* (Mravinac et al. 2004).

Heterochromatinization and diverse silencing mechanisms, for example, PEV in *Drosophila*, may be triggered by mechanisms that recognize unusual structures of inverted repeats within satellite monomers (Dernburg and Karpen 2002). Insertions in direct or reverse orientation are mostly products of recombination, a major force that forms new satellite monomers by sequence rearrangements (Dover 1986, 2002). Inverted or direct repeats in satellite monomers might also be remnants of transposable element-like

genomic components and might indicate transposition as a mechanism of satellite spread (Dover 2002).

The impact of nucleotide sequence and its structural characteristics on possible functional properties of satellite DNAs is still poorly understood. In particular, there is a lack of data concerning evolution and putative functional constraints on satellite monomers built of inverted sequence segments. Here we characterize a complex satellite DNA from the beetle *Tribolium brevicornis* and propose the evolutionary pathway of its monomer sequence. The repeating unit is built of the two inversely oriented subunits that form the longest inverted repeat ever described in any satellite monomer. Mutational profiles of inversely oriented subunits are compared in order to deduce evolutionary dynamics of elements of the potential dyad structure. In addition, putative functional nucleotide sequence elements are detected and compared to those found in other *Tribolium* satellites.

## Materials and Methods

### *Insects*

*Tribolium brevicornis* was purchased from the Central Science Laboratory (Sand Hutton, York, UK), and kept as a laboratory culture.

### *DNA Extraction, Cloning, and Sequencing*

Genomic DNA was extracted from adults and larvae by standard phenol extraction and ethanol precipitation protocol. Digestions with restriction endonucleases were performed according to the manufacturer's specifications (Roche). Restriction fragments were electrophoretically separated on a 1% agarose gel, and fragments of interest cut out and purified with the Qiaquick Gel Extraction Kit (Qiagen) and subsequently ligated in the pUC18-*Sma*I plasmid vector. Positives were detected by hybridization at 65°C (expected homology > 80%), with a portion of eluted fragments labeled with digoxigenin by the random priming method, using the DIG DNA Labeling and Detection Kit (Roche). Recombinant clones were sequenced by the MWG-Biotech sequencing service (Ebensberg, Germany).

### *Genomic Southern and Dot-Blot Quantification*

For Southern analysis, 5  $\mu$ g of genomic DNA was digested with restriction endonucleases, electrophoresed on a 1% agarose gel, and blotted onto a positively charged nylon membrane (Roche). To prepare the probe, cloned satellite monomer was excised from the plasmid, purified, and labeled using the DIG DNA Labeling and Detection Kit (Roche). Membranes were hybridized under high-stringency conditions, at 68°C (homology > 90% allowed). Relative genomic contribution of the satellite DNA was determined by densitometry of hybridization signals obtained from serially diluted genomic DNA spotted onto a nylon membrane. The obtained signals were compared to the calibration curve constructed from hybridization signals of diluted cloned satellite monomers. Hybridization was performed with labeled satellite monomer under the same conditions as for Southern analysis.

## Sequence Analysis

Multiple-sequence alignments were performed using the default parameters of the ClustalX v. 1.81 program (Thompson et al. 1997). Pairwise sequence divergences were calculated according to the best-fit model of nucleotide evolution HKY85 + G (Hasegawa–Kishino–Yano with gamma-distribution [Hasegawa et al. 1985]), selected by the hierarchical likelihood ratio test (hLRT) in Modeltest 3.06 (Posada and Crandall 1998). A distance tree was built by the neighbor-joining (NJ) method using the PAUP\* v. 4b10 computer package (Swofford 1998). A parsimony tree was obtained after 1000 random addition searches with TBR branch swapping and under the Multrees option, using the same computer package. All nucleotide positions in the matrix, including gap positions, were considered equivalent and weighted equally. The strict consensus tree of the 20 most parsimonious trees is presented. Bootstrap values were calculated based on 1000 replicates.

Motifs homologous to the human CENP-B box (Masumoto et al. 1998) as well as direct and inverted repeats were detected using the MicroGenie (Beckmann) package and MegAlign program (Dnastar). DNA polymorphism and putative gene conversion tracts were estimated using the DnaSP program, v. 3.99 (Rozas and Rozas 1999). The conserved and variable segments in the satellite DNA sequence were defined by sliding window analysis implemented in the same program, using a window size of 10 bp and a step size of 1, as described by Mravinac et al. (2004). Frequency of A + T tracts (O value) was determined with COMPSEQ (<http://www.bioweb.pasteur.fr/seganal/interfaces/compseq.html>) and the expected frequency in a random sequence of the same A + T content (E values) was calculated according to Barceló et al. (1998). The deviation of the observed from the expected frequency was determined using chi-square statistics. GenBank searches were done using the program BLAST (<http://www.ncbi.nlm.nih.gov/blast/>).

The evolutionary trends at each nucleotide position of satellite DNA were determined by the method of Strachan et al. (1985). According to this analysis variation at nucleotide positions between the two groups of sequences is classified in six categories, representing transition stages of concerted evolution (see Fig. 4 in Strachan et al. 1985). Class 1 represents homogeneity of all clones at a given position within two sets of sequences, class 5 represents mutations fixed in all clones of each of the sets, while classes 2–4 are intermediate stages. Class 6 represents introduction of subsequent mutations in a homogenized set and, thus, indicates the start of a new cycle of replacements.

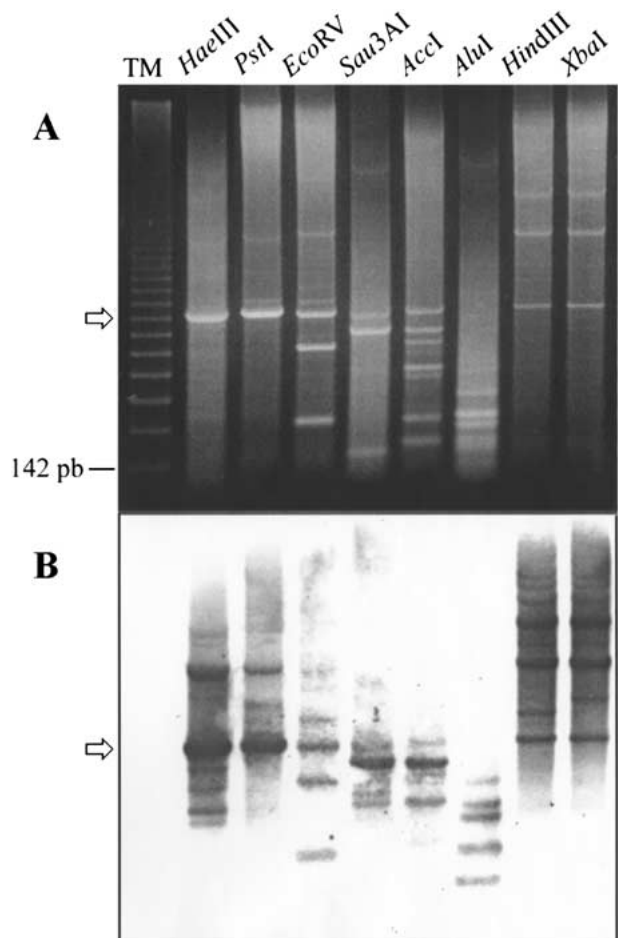
## Fluorescent In Situ Hybridization (FISH)

Chromosome spreads were obtained from male gonads by the standard squash method and FISH was done as described earlier (Ugarković et al. 1996a). The hybridization signal was detected by the avidin–fluorescein isothiocyanate (FITC) system (Vector, USA) with amplification. The chromosomes were counterstained with fluorochrome DAPI and photographed in an Opton microscope.

## Results

### Detection and Cloning of Satellite DNA in the Beetle *Tribolium brevicornis*

Electrophoretic separation of *T. brevicornis* genomic DNA digested with a series of restriction enzymes revealed bands typical for highly abundant repetitive sequences (Fig. 1A). Bands of approximately 1100



**Fig. 1.** Detection of *T. brevicornis* satellite DNA. **A** Agarose gel electrophoresis of genomic DNA restricted with enzymes indicated above each lane. Partially digested *Tenebrio molitor* satellite DNA is used as a size marker (lane TM). The arrow indicates the position of a band of approximately 1100 bp. **B** Southern blot of the gel shown in A after hybridization with labeled TBREV monomers, cloned from the band obtained after *HaeIII* digestion of genomic DNA.

bp, obtained after restrictions with *HaeIII* and *EcoRV*, were cut out from the gel, and DNA was extracted and cloned into a plasmid vector. Positive clones were detected after hybridization with a labeled fraction of material extracted from the *HaeIII* band. The mixture of inserts obtained from positive clones was labeled and used as a probe in Southern hybridization analysis (Fig. 1B). Multimers of basic length, visible up to the hexamer after digestion with *HindIII* and *XbaI*, are characteristic for tandemly repeated sequences. Intermediate bands, visible in all lanes, can be generated by additional recognition sites existing for some restriction endonucleases (see below).

Copy number of TBREV satellite DNA was determined after dot-blot hybridization of serial dilutions of genomic DNA, using a mixture of cloned satellite monomers as a probe (not shown). The highly abundant *T. brevicornis* satellite makes 21.2% of the



Fig. 2. Alignment of nucleotide sequences obtained by random cloning of TBREV satellite monomer and from them derived consensus sequence according to the majority rule. Identity to the consensus is indicated by dots; dashes denote indels. Tracts of nucleotides A or T  $\geq 3$  are in **boldface**, and motifs similar to the CENP-B box are in lowercase *italics*. Sequence segments determined to be variable in sliding window analysis are shadowed, while those determined to be conserved are underlined. Segment of the cloned monomer TBREV137 from position 884 to the end can be unambiguously aligned in reverse orientation to the first 101 nucleotides of all other monomers. This particular segment is presented in the alignment as TBREV 137r. Sequences TBREV111, 137, and 137r are separated from the rest because they were not included in phylogenetic analyses (Figs. 4 and 5). EMBL/GenBank accession numbers of cloned TBREV satellite monomers are AY672602–AY672617.

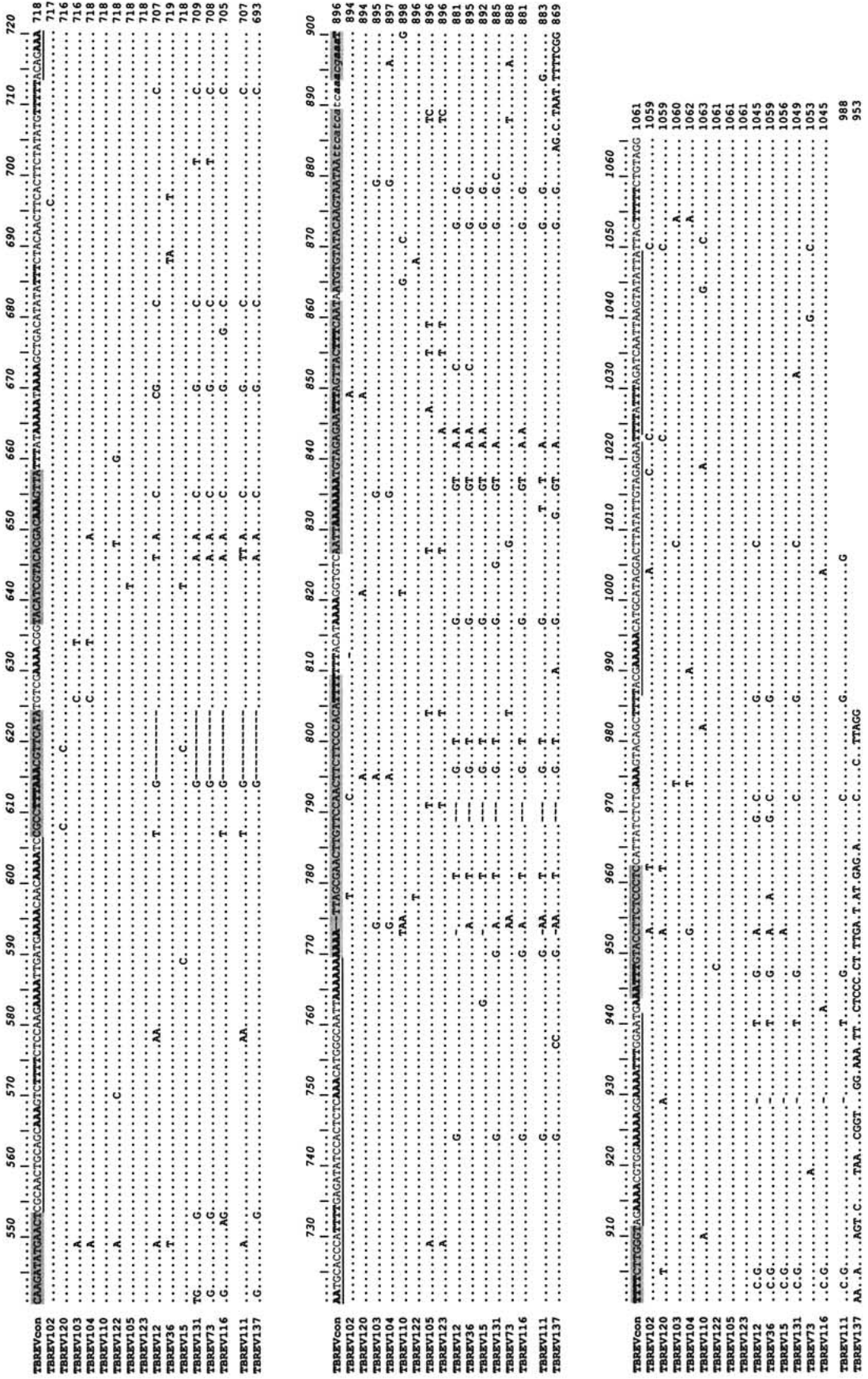


Fig. 2. Continued.

genome. According to the genome size of  $3.540 \times 10^8$  bp determined by Alvarez-Fuster et al. (1991), this is  $7.07 \times 10^4$  copies per haploid complement.

### Sequence Analysis of the TBREV Satellite

The sequence analysis of 16 randomly cloned fragments of the *T. brevicornis* satellite (TBREV), 15 of them from the *Hae*III band and 1 from the *Eco*RV band, supported observed migration of electrophoretic fragments (Fig. 1). Only a single restriction site was detected in the consensus monomer sequence for endonucleases *Hae*III and *Pst*I, several restriction sites exist for *Eco*RV, *Sau*3AI, *Acc*I, and *Alu*I, and *Hind*III and *Xba*I do not cut within the satellite. However, restriction sites for the latter two can be generated due to single point mutations, and therefore digestions of genomic DNA with these enzymes produce a characteristic ladder of satellite multimers. The alignment revealed a 1061-bp-long consensus sequence (Fig. 2). Since the fragment TBREV73 originated from the *Eco*RV band, it was aligned to *Hae*III-cloned fragments after circular permutation of its sequence. However, it did not show any deviation from other clones. This is consistent with a regular tandem arrangement of repetitive units. A schematic presentation of cloned monomers and their subunit organization is shown in Fig. 3.

Two cloned fragments differ from the rest in the analyzed set: TBREV111 and TBREV137 are 73 and 108 nucleotides shorter, respectively. Observed trimming occurred during restriction digestion in the cloning procedure as a consequence of additional *Hae*III restriction sites envisaged on indicated positions of the two monomer variants (Fig. 2). The alignment pattern of TBREV137 deserves particular attention. Its nucleotide sequence follows the consensus satellite sequence until position 883, then it reverses the orientation and follows the consensus sequence from position 1 to position 101 (Figs. 2 and 3). To avoid comparisons with truncated and/or rearranged monomer sequences, these two clones were excluded from further analyses.

Sequence comparison showed high homogeneity of aligned monomers (Fig. 2), with variable sites distributed on 212 positions. In total there are 232 changes: 204 (88%) of them are single-nucleotide substitutions, and 28 (12%) changes are insertion or deletion events (indels). Of 232 changes, 94 (41%) are detected at a particular site of a single clone (autapomorphies), while 138 (59%) are shared within a subset of monomer variants (synapomorphies). Ratio of transitions to transversions is 1.78.

The nucleotide diversity (per site),  $P_i = 0.04988$ , was calculated by the DnaSP program (Rozas and

Rozas 1999) with gaps excluded from the calculation. Standard deviation is 0.00680. Distribution of variable sites along the sequence was determined by sliding window analysis, with a window length = 10 and a step size = 1. Considered significant as variable or as conserved are those windows with a diversity  $\geq$  (average + 2 SD) or  $\leq$  (average - 2 SD), respectively. This analysis revealed 14 variable and 13 conserved regions varying in size from 16 to 69 bp and interspersed within a 1061-bp-long monomer (Fig. 2).

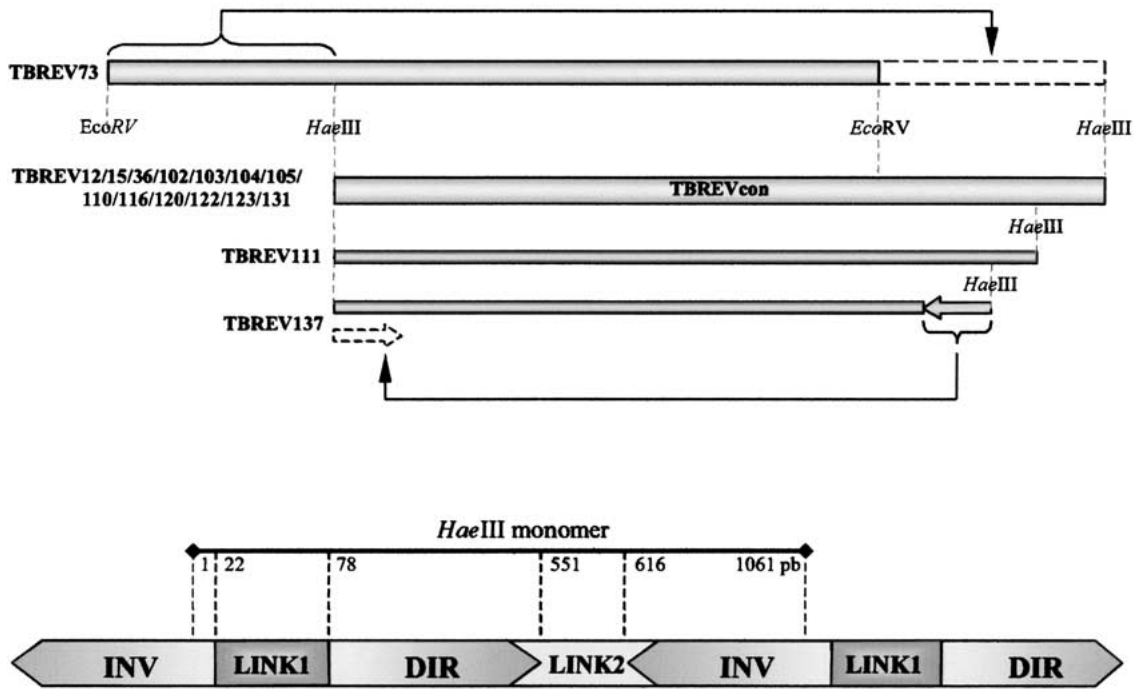
The NJ tree of 14 satellite monomer sequences was constructed according to the HKY+G model (Fig. 4). Differentiation of sequence groups in the dendrogram is due to multiple nucleotide substitutions shared by subsets of monomers. Sequences in a branch marked with an asterisk are characterized by stretches of diagnostic nucleotides. For TBREV12, 15 and 36 diagnostic nucleotides are localized within a segment of 125 nt, from position 781 to position 905 of the consensus (Fig. 2). Another subgroup is defined by TBREV73, 116 and 131, which can be further differentiated by a number of exclusive changes distributed along the monomer, except from position 906 to the end of the consensus sequence for TBREV116 and 131, and from position 781 to the end for TBREV73. A complex pattern of stretches of shared mutations indicates that more than one conversion event homogenized the satellite sequence at submonomer length.

The content of A+T nucleotides in the consensus sequence is 69.2%. The high frequency of A+T nucleotides results in many tracts of A or T  $\geq 3$  which show an ordered arrangement along the sequence, according to the comparison with a random sequence of the same length and composition (Fig. 2). Expected frequency of A or T  $\geq 3$  tracts (E) in a random sequence is 135, while the frequency observed in the TBREV satellite is  $O = 281$ . The difference between expected and observed values is statistically significant,  $P < 0.0001$ .

No sequence homology to any other known sequence could be detected in a search of GenBank in July 2004. In addition, no open reading frames of significant length could be detected in the TBREV satellite sequence.

### Subrepeat Organization of the Satellite Monomer

Internal organization of TBREV satellite monomer shows two prominent sequence segments repeated in an inverted orientation within the monomer (Fig. 3). Inverted subunits of 473 and 467 bp, named DIR and INV, respectively, share a similarity of 82% and have a high potential to form an energetically stable dyad structure. The subunits are separated by two linker



**Fig. 3.** Schematic presentation of cloned satellite monomers and their subunit organization. Restriction sites and clone names are indicated. Clone TBREV73 should be circularly permuted in order to be aligned with other clones. Inverted segment of the clone TBREV137 is indicated by an arrow; the dashed arrow shows the

position of this segment in the sequence alignment. Organization of inversely oriented subunits DIR and INV and unique sequence elements LINK1 and LINK2 in the satellite sequence is presented along with the corresponding positions in the *HaeIII* monomer.

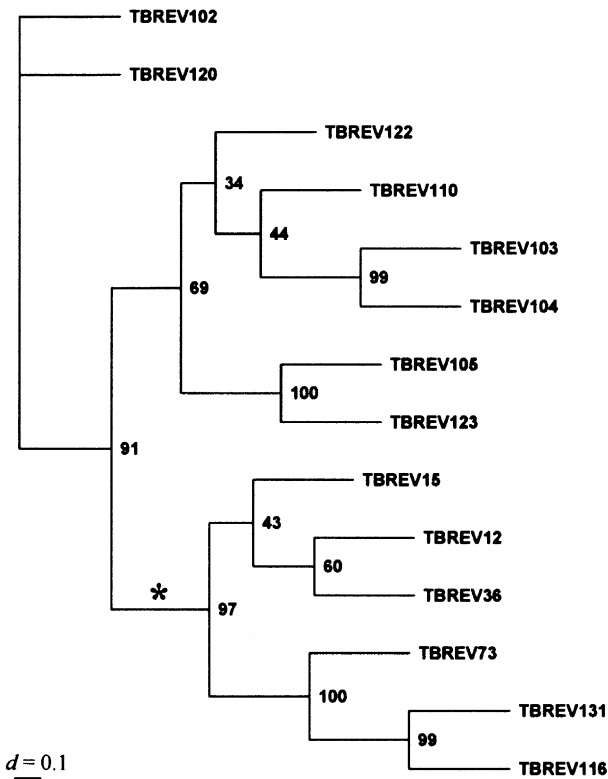
segments, LINK1 of 56 bp and LINK2 of 65 bp, unique in their nucleotide composition.

Phylogenetic analysis separated subunits into two major clades corresponding to DIR and INV sequences (Fig. 5). Although DIR and INV sequences are clearly separated, the topology of the tree presented in Fig. 5 is not symmetrical, because DIR and INV subunits originating from the same set of satellite monomers do not always group in the same arrangement. For example, DIR103 and DIR104 are closely related in the tree, and the same is true for INV103 and INV104. This distribution is expected for the evolution of satellite monomers by proportional accumulation of mutations along the monomer sequence, and in that case the topology of DIR and INV clades should be symmetrical. Contrarily, while INV102 and INV120 form a closely related pair in the tree, DIR102 is associated with DIR12, and DIR120 is placed in a distant branch. This discordance can be explained by the spreading of diagnostic changes and homogenization of monomer variants on the level of subunit or near-subunit size.

Alignment of consensus sequences of DIR and INV subunits is shown in Fig. 6. Variable and conserved segments determined for each group of sequences are indicated on the corresponding consensus. In general, the distribution of variable and

conserved segments between DIR and INV subunits coincides only partially. For example, although the long stretch of 38 nucleotides limited by asterisks in Fig. 6 is devoid of any nucleotide substitutions that would differentiate between the two consensus sequences, the same segment is categorized as conserved in the DIR group of subunits and as variable in the INV group. Observed discrepancy might indicate that constraints on the sequence evolution can be orientation-dependent or are altered due to already accumulated sequence divergence between DIR and INV subunits.

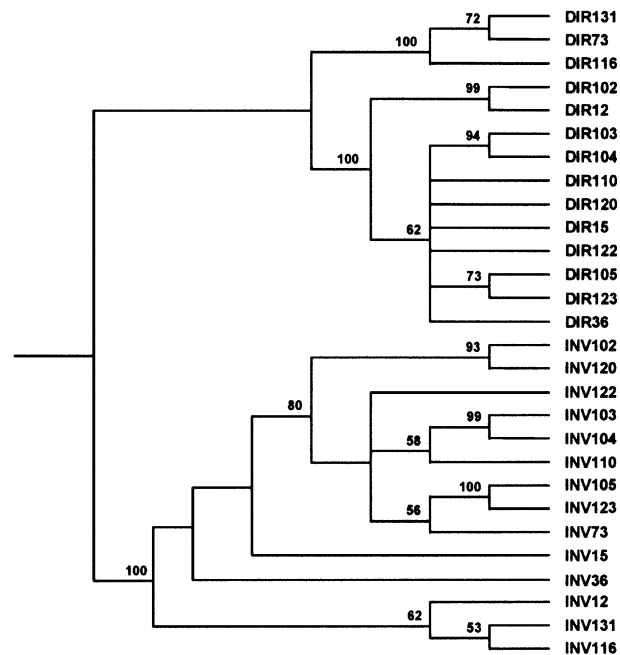
In addition to the two inverted subunits, numerous short direct and inverted repeats up to ~20 bp can be detected along the monomer sequence (data not shown). They are repeated only two or three times within the monomer in an irregular and sometimes overlapping manner which does not allow detection of any further organizational patterns. The most prominent inverted repeats, one of them located immediately near the tract composed of exclusively A+T nucleotides, are presented in Fig. 6. Sequence alignment revealed nucleotide substitutions that have accumulated between the two subunits, but no short direct and inverted repeats could be found to be specific to any of them. Accordingly, they all existed before emergence of inversely oriented subunits and formation of the current satellite monomer.



**Fig. 4.** The unrooted dendrogram of the relationships among 14 monomeric units of the TBREV satellite obtained by the NJ method. Bar represents the genetic distance  $d = 0.1$  of pairwise comparisons calculated according to the HKY85+G model. Numbers at each node indicate the percentage of trees representing the particular node of 1000 bootstrap replicates. An asterisk marks the cluster of monomers characterized by stretches of diagnostic nucleotides (Fig. 2 and explanation in the text).

#### *Transition Stages in the Evolution of the TBREV Satellite*

Transition stages during the spread of new mutations in TBREV satellite were examined using the analysis by Strachan et al. (1985), recently widely implemented in analyses of tandemly repeated sequences (Pons et al. 2002; Lorite et al. 2004). First, patterns of variation at individual nucleotide positions were classified in the monomer sequence as a whole, as well as in particular segments DIR, INV, LINK1, and LINK2 (Table 1). Two distinct branches of the NJ tree shown in Fig. 4 were defined as two groups of sequences in the comparison. Since monomers TBREV102 and 120 did not lean to any group, they were excluded from this analysis. Changes that could be categorized as classes 5 and 6 were not detected, while nucleotide positions classified in stages 1 and 2 represent over 90% of all informative sites. All other informative sites in LINK1 segment and all but one in the DIR subunit belong to class 3, while in LINK2 and INV they mostly belong to class 4. This pattern reveals an



**Fig. 5.** The midpoint rooted strict consensus tree of 20 most parsimonious trees of DIR and INV subunits. The interpolated bootstrap values represent the percentages of trees supporting the particular node of 1000 bootstrap replicates. Bootstrap values below 50% are not indicated. Clustering of DIR and INV subunits into two well-defined clades was supported by the NJ tree as well (not shown).

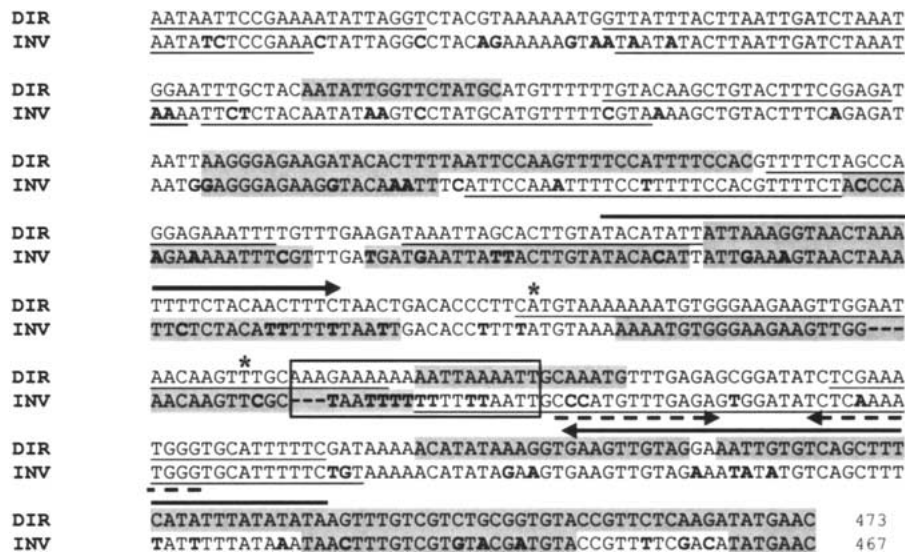
early stage in differentiation of two groups of satellite monomers in formation of subfamilies.

The same analysis was used to categorize patterns of mutations between sequences of inversely oriented subunits DIR and INV (Table 2). The occurrence of positions categorized in class 5 or 6 (DIR vs. INV in Table 2) shows advanced divergence of DIR and INV subunits. Besides the total number of nucleotides in each transition stage, the fraction of changed nucleotides is calculated for each of the DIR or INV group of subunits separately. This cannot be done for classes 1 and 5, which are by definition completely homogenized in both groups of sequences. A higher number of positions of class 6 are found among INV subunits, while the number of changes in other classes is balanced for both groups.

#### *CENP-B Box-like Motifs in the TBREV Satellite*

Motifs similar to the human CENP-B box (Masumoto et al. 1989) were detected on three positions of the monomer sequence, on two of them with 71% and on one with 65% similarity (Fig. 2). However, the three CENP-B box-like sequences are irregularly distributed on the satellite monomer and they appear in both orientations and lack significant similarity among themselves. These facts bring into question the plausibility of potential CENP-B boxes, although the detected similarity exceeds similarities that most





**Fig. 6.** Alignment of consensus sequences of DIR and INV subunits. Sequence divergences between them are indicated in **boldface** at INV consensus sequence. Arrows mark the most remarkable inverted repeats. A block of pure A and T nucleotides is boxed. Localization of conserved and variable segments (underlined and shaded, respectively) according to sliding window analysis is indicated along each subunit consensus sequence.

**Table 1.** Percentage of nucleotide positions in the TBREV monomer and its segments, distributed into six classes of transition stages according to Strachan et al. (1985)

	TBREV	DIR	INV	LINK 1	LINK 2
Informative position <sup>a</sup>	97.28	97.26	96.80	98.21	100
Class 1	83.40	83.33	82.60	89.09	84.62
Class 2	9.94	10.39	10.97	1.82	7.69
Class 3	3.96	6.06	1.54	9.09	1.54
Class 4	2.70	0.22	5.07	0	6.15
Class 5	0	0	0	0	0
Class 6	0	0	0	0	0

<sup>a</sup>Percentage of total number of nucleotide positions.

**Table 2.** Percentage of nucleotide positions between compared DIR and INV subunits, falling into six classes of transition stages according to Strachan et al. (1985)

	DIR vs. INV	DIR <sup>a</sup>	INV <sup>b</sup>
Informative positions <sup>c</sup>	92.42		
Class 1	62.87	—	—
Class 2	20.96	50	50
Class 3	0	0	0
Class 4	6.38	54	46
Class 5	7.06	—	—
Class 6	2.73	25	75

<sup>a</sup>Percentage of changes in DIR subunits (corresponding nucleotide positions in INV subunits are completely homogenized).

<sup>b</sup>Percentage of changes in INV subunits (corresponding nucleotide positions in DIR subunits are completely homogenized).

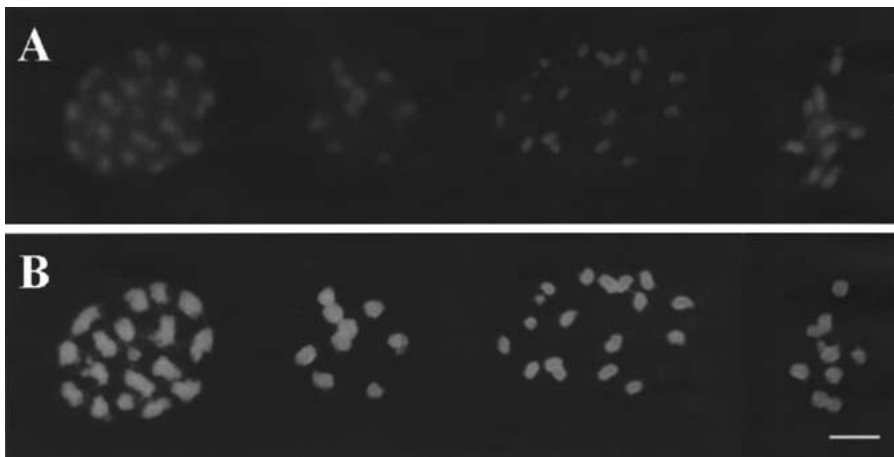
<sup>c</sup>Percentage of total number of nucleotide positions.

of the published CENP-B-like boxes share with the human CENP-B box (Lorite et al. 2002 and references therein).

#### Chromosomal Localization of the TBREV Satellite Sequence

*T. brevicornis* has  $2n = 18$  chromosomes (Moore and Sokoloff 1982). Large blocks of pericentromeric

heterochromatin are visible on all chromosomes of the complement after staining with the fluorochrome DAPI (Fig. 7A). FISH with labeled TBREV satellite DNA revealed strong hybridization signal overlapping with pericentromeric regions on all chromosomes (Fig. 7B). This distribution of pericentromeric satellites, irrespective of the copy number, is characteristic for all tenebrionid species examined until now (e.g., Durajlija-Žinić et al. 2000; Meštrović et al. 2000; Mravinac et al. 2004).



**Fig. 7.** FISH of *T. brevicornis* chromosomes ( $2n = 18$ ). **A** Chromosomes stained with DAPI. **B** After hybridization with FITC-labeled cloned TBREV satellite monomer. Bar represents 10  $\mu\text{m}$ .

## Discussion

We propose a scenario that could be projected for the evolution of the TBREV satellite in which directly and inversely duplicated short sequence segments first contributed to the formation of an initial  $\sim 470$ -bp-long element. In the following step, the two inversely duplicated copies of this element were amplified together and became subunits in a new complex repeating unit which formed tandem repeats of the extant TBREV satellite. When complex monomers are composed of former units homogenized in a higher-order register, sequence homogeneity is always higher among whole complex repeats than among their constitutive subunits (Warburton and Willard 1990; Willard and Waye 1987). However, it is not possible to discern whether 18% of sequence divergence between the TBREV satellite subunits already existed in the moment of amplification in a higher-order register or accumulated subsequently. Two unrelated spacer segments that alternate between  $\sim 470$ -bp repeats might simply represent sequences that were resident in the site of the duplication/inversion event (Fig. 3). Presumably, a new complex repeating unit spread through the genome in the process of molecular drive (Dover 1986, 2002) and became the dominant satellite DNA component in the *T. brevicornis* genome. During that process, it probably replaced some ancient form of satellite sequence that had already existed in the genome, as proposed in a model of satellite life history (Nijman and Lenstra 2001).

A trend toward increasing repeat length and complexity, first by forming a monomer from short sequence motifs and then by merging two or more monomers in a new higher-order repeating (HOR) unit, has been observed in some satellite DNAs. Theoretical models that predict low ratios of unequal crossing-over to mutation, presumed to be typical for heterochromatic regions, are consistent with the for-

mation of longer and more complex repetitive units (Stephan and Cho 1994). For example, this organizational feature has been detected in satellites from the cave beetle *Pholeuon proserpine glaciale* (Pons et al. 2003) and Bovidae species (Lee et al. 1997; Modi et al. 2004). In the beetle *Palorus subdepressus* the complex satellite monomer is formed by two adjacent units, which are composed of an octanucleotide alternating with elements of an inverted repeat (Pohl et al. 1998).

The inversion of whole subunits in the formation of HOR is not a frequent feature in the evolution of complex satellite monomers. It has been described in the satellite of *Drosophila simulans*, which is composed of a 5-bp-long motif repeated three times, one copy being reversely oriented (Lohe and Roberts 1988). In the parasitoid wasp *Trichogramma brassicae* the satellite DNA evolved through duplication and inversion of an 80-nt-long subunit and insertion of a partially duplicated sequence element (Landais et al. 2000). Similarly, duplication and inversion of repeating units plus insertion of unrelated sequence and homogenization in higher-order register formed complex monomers in satellite DNAs of the beetle *Tribolium madens* (Ugarković et al. 1996b) and species of the genus *Pimelia* (Pons et al. 2004). The finding of a recombinant clone with inverted fragment of an adjacent satellite monomer supports proneness to inversion of the TBREV satellite sequence. Although further characterization has not been done, it can be assumed that sites like this could represent a source of new repeating units that can harbor even longer inverted repeats.

Analysis of phylogenetic distribution of DIR and INV subunits with respect to their localization in cloned monomer variants indicates that, simultaneously with the spread of the whole TBREV monomers, homogenization is effective on the lower scale as well, spreading fragments of subunit-similar

lengths. This process results in disunited pairs of adjacent DIR and INV subunits, randomizing their distribution in satellite monomers. Formation of satellite subfamilies in the mollusk *Donax trunculus* and occurrence of stretches of mutations shared by some variants have been explained as a concurrent action of gene conversion on fragments of the monomer length as well as on short tracts that can be only a few base pairs long (Plohl and Cornudella 1996). According to the distribution of shared mutations, TBREV monomers or their subunits form clusters of variants in analyses presented in Figs. 4 and 5, thus indicating a trend toward formation of subfamilies. Formation of subfamilies is not characteristic for tenebrionid satellites, in which efficient homogenization within and between chromosomes causes random distribution of mutations and lack of any differentiation of monomer variants, such as in the 142-bp-long satellite from the mealworm beetle *Tenebrio molitor* (Plohl et al. 1992). The observed grouping of TBREV monomers could be a consequence of their extraordinary length and complex structure, which might impair the efficacy of homogenization mechanisms.

The distribution of mutations in the advanced stage of the homogenization process, characterized by subsequent replacements of already fixed group-specific differences, indicates that the INV subunit introduces new mutations more frequently than the DIR subunit. The observed tendency towards increased divergence between DIR and INV subunits does not favor existence of constraints in order to preserve the possible dyad structure that might be formed by these sequence elements.

Despite uncommon length and complex architecture of the monomeric unit, and lack of any sequence similarity, the TBREV satellite retained all structural features of centromeric satellites detected in other *Tribolium* species (Juan et al. 1993; Plohl et al. 1993; Ugarković et al. 1996a, b; Mravinac et al. 2004). It has been proposed that an inverted repeat in the vicinity of an A+T tract, clustering of A and T nucleotides, CENP-B box-like motifs, and alternating tracts of conserved and variable sequence segments might represent functional elements relevant for specific protein recognition and condensation of heterochromatin (Mravinac et al. 2004 and references therein). Due to the sequence divergence, “small” dyad structures formed by short inverted repeats are not equally probable at equivalent positions within the DIR and INV subunits of the TBREV satellite. Lack of sequence conservation and lower probability of dyad structure formation in one of the subunits do not necessarily mean lack of functionality. For example, only a subset of  $\alpha$  satellite 17-bp sequence motifs is able to bind CENP-B protein, since this motif is located in a variable region of the satellite

monomer (Hall et al. 2003), while a higher-order arrangement of  $\alpha$  satellite repeats enables phased distribution of functional CENP-B binding sites (Ikeno et al. 1994).

Without extensive additional experimental research it is difficult to evaluate the probability of *in vivo* formation of stable dyad structures in highly compact centromeric heterochromatin and determine their possible functional importance. The short inversely oriented sequence motif in the vicinity of an A+T tract found in TBREV and in all examined *Tribolium* satellites (Ugarković et al. 1996a; Mravinac et al. 2004) might represent a signal for protein binding per se, simply due to the arrangement of DNA sequence elements, even without formation of secondary structures. Alternatively, long inverted subunits may be evolutionarily favored because they can form far more stable dyad structure than short inverted repeats and might adopt the function of these repeats in protein recognition and heterochromatin condensation. It is possible that due to the high stability of long secondary structure, the accumulation of differences in subunits of TBREV satellite can be tolerated until they do not disturb its integrity significantly.

Directly and inversely repeated sequences are common features of many transposable elements, and transposition is considered one of the mechanisms of nonreciprocal transfer in the process of concerted evolution (Dover 2002). Transposable elements or their segments can be amplified in tandem and form centromeric satellites, as has been reported in *Drosophila* (Heikkinen et al. 1995) and in whales (Kapitonov et al. 1998). Inverted repeats described recently in the satellite of the darkling beetle *Misolampus goudoti* (Pons 2004) resemble structures found in miniature inverted-repeat transposable elements (MITEs) detected in the teleost fish (Izsvak et al. 1999). A repetitive element similar to MITE is abundant in the closely related *D. obscura* species group, and its derivative forms abundant satellite DNA in *D. guanche* (Miller et al. 2000). We can speculate that the organizational features of TBREV satellite based on inversely repeated subunits might be reminiscent of some form of transposable element from which it originated and that transposition might have contributed to its formation and spread.

In conclusion, characterization of the abundant pericentromeric satellite in *T. brevicornis* revealed a repeating unit of complex structure and evolutionary history, with the longest inversely oriented subunits detected in any satellite until now. However, as the nucleotide sequence in inversely oriented subunits evolves divergently, it does not seem to be under constraints to maintain the secondary structure. The presence of structural features shared with other

known *Tribolium* centromeric satellites indicates that, although unrelated in sequence and monomer organization, all of them might contribute to some similar function.

**Acknowledgments.** This work was supported by the Research Fund of Republic of Croatia, project Nos. P0098075 to M. Plohl, P0098074 to Đ. Ugarković at Ruder Bošković Institute, and P0119124 to M. Kalafatić at University of Zagreb, Faculty of Science.

## References

- Alvarez-Fuster A, Juan C, Petitpierre E (1991) Genome size in *Tribolium* flour-beetles: inter- and intraspecific variation. *Genet Res* 58:1–5
- Barceló F, Gutierrez F, Barjau I, Portugal J (1998) A theoretical perusal of the satellite DNA curvature in tenebrionid beetles. *J Biomol Struct Dyn* 16:41–50
- Bigot Y, Hamelin M-H, Periquet G (1990) Heterochromatin condensation and evolution of unique satellite-DNA families in two parasitic wasp species: *Diadromus pulchellus* and *Eupelmus vvillei* (Hymenoptera). *Mol Biol Evol* 7:351–364
- Castagnone-Sereno P, Leroy F, Abad P (2000) Cloning and characterization of an extremely conserved satellite DNA family from the root-knot nematode *Meloidogyne arenaria*. *Genome* 43:346–353
- Cheng Z, Dong F, Langdon T, Ouyang S, Buell CR, Gu M, Blattner FR, Jiang J (2002) Functional rice centromeres are marked by a satellite repeat and a centromere specific retrotransposon. *Plant Cell* 14:1691–1704
- Dernburg AF, Karpen GH (2002) A chromosome RNAissance. *Cell* 111:159–162
- Dover GA (1986) Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. *Trends Genet* 2:159–165
- Dover GA (2002) Molecular drive. *Trends Genet* 18:587–589
- Durajlija-Zinić S, Ugarković Đ, Cornudella L, Plohl M (2000) A novel interspersed type of organization of satellite DNAs in *Tribolium maodens* heterochromatin. *Chromosome Res* 8:201–212
- Hall SE, Kettler G, Preuss D (2003) Centromere satellites from *Arabidopsis* populations: Maintenance of conserved and variable domains. *Genome Res* 13:195–205
- Hasegawa M, Kishino K, Yano T (1985) Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 22:160–174
- Heikkinen E, Launonen V, Müller E, Bachmann L (1995) The *pvB370* satellite DNA family of the *Drosophila virilis* group and its evolutionary relation to mobile dispersed genetic *pDv* elements. *J Mol Evol* 41:604–614
- Ikeno M, Masumoto M, Okazaki T (1994) Distribution of CENP-B boxes reflected in CREST centromere antigenic sites on long-range  $\alpha$ -satellite DNA arrays of human Chromosome 21. *Hum Mol Genet* 3:1245–1257
- Izsvák Z, Ivics Z, Shimoda N, Mohn D, Okamoto H, Hackett PB (1999) Short inverted-repeat transposable elements in teleost fish and implications for a mechanism of their amplification. *J Mol Evol* 48:13–21
- Juan C, Vazquez P, Rubio JM, Petitpierre E, Hewitt GM (1993) Presence of highly repetitive DNA sequences in *Tribolium* flour-beetles. *Heredity* 70:1–8
- Kapitonov VV, Holmquist GH, Jurka J (1998) L1 repeat is a basic unit of heterochromatin satellites in Cetaceans. *Mol Biol Evol* 15:611–612
- Landais I, Chavigny P, Castagnone C, Pizzol J, Abad P, Vanlerberghe-Masutti F (2000) Characterization of a highly conserved satellite DNA from the parasitoid wasp *Trichogramma brassicae*. *Gene* 255:65–73
- Lee C, Court DR, Cho C, Haslett JL, Lin C-C (1997) Higher-order organization of subrepeats and the evolution of cervid satellite I DNA. *J Mol Evol* 44:327–335
- Lohe AR, Roberts PA (1988) Evolution of satellite DNA sequences in *Drosophila*. In: Verma RS (ed) *Heterochromatin*. Cambridge University Press, Cambridge, MA, pp 148–186
- Lorite P, Carrillo JA, Tinaut A, Palomeque T (2002) Comparative study of satellite DNA in ants of the *Messor* genus. *Gene* 297:113–122
- Lorite P, Carrillo JA, Tinaut A, Palomeque T (2004) Evolutionary dynamics of satellite DNA in species of the genus *Formica* (Hymenoptera, Formicidae). *Gene* 332:159–168
- Malik HS, Henikoff S (2001) Adaptive evolution of Cid, a centromere-specific histone in *Drosophila*. *Genetics* 157:1293–1298
- Martienssen RA (2003) Maintenance of heterochromatin by RNA interference of tandem repeats. *Nat Genet* 35:213–214
- Masumoto H, Masukata H, Muro Y, Nozaki N, Okazaki T (1989) A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromere satellite. *J Cell Biol* 109:1963–1973
- Meštrović N, Mravinac B, Juan C, Ugarković Đ, Plohl M (2000) Comparative study of satellite sequences and phylogeny of five species from the genus *Palorus* (Insecta, Coleoptera). *Genome* 43:776–785
- Miller WJ, Nagel A, Bachmann J, Bachmann L (2000) Evolutionary dynamics of the *SGM* transposon family in the *Drosophila obscura* species group. *Mol Biol Evol* 17:1597–1609
- Modi WS, Ivanov S, Gallagher DS (2004) Concerted evolution and higher-order repeat structure of the 1.709 (satellite IV) family in bovids. *J Mol Evol* 58:460–465
- Moore RD, Sokoloff A (1982) The determination of chromosome number in *Tribolium brevicornis* and *Tribolium anaphe*. *Trib Info Bull* 22:116–117
- Mravinac B, Plohl M, Ugarković Đ (2004) Conserved patterns in the evolution of *Tribolium* satellite DNAs. *Gene* 332:169–177
- Nijman IJ, Lenstra JA (2001) Mutation and recombination in cattle satellite DNA: A feedback model for the evolution of satellite DNA repeats. *J Mol Evol* 52:361–371
- Plohl M, Cornudella L (1996) Characterization of a complex satellite DNA in the mollusk *Donax trunculus*: analysis of sequence variation and divergence. *Gene* 169:157–164
- Plohl M, Borštnik B, Lucijanić-Justić V, Ugarković Đ (1992) Evidence for random distribution of sequence variants in *Tenebrio molitor* satellite DNA. *Genet Res* 60:7–13
- Plohl M, Lucijanić-Justić V, Ugarković Đ, Petitpierre E, Juan C (1993) Satellite DNA and heterochromatin of the flour beetle *Tribolium confusum*. *Genome* 36:467–475
- Plohl M, Meštrović N, Bruvo B, Ugarković Đ (1998) Similarity of structural features and evolution of satellite DNAs from *Palorus subdepressus* (Coleoptera) and related species. *J Mol Evol* 46:234–239
- Pons J (2004) Cloning and characterization of a transposable-like repeat in the heterochromatin of the darkling beetle *Misolampus goudoti*. *Genome* 47:1–6
- Pons J, Petitpierre E, Juan C (2002) Evolutionary dynamics of satellite DNA family PIM357 in species of the genus *Pimelia* (Tenebrionidae, Coleoptera). *Mol Biol Evol* 19:1329–1340
- Pons J, Bucur R, Vogler AP (2003) Higher-order repeats in the satellite DNA of the cave beetle *Pholeuon proserpinae glaciale* (Coleoptera: Cholevidae). *Hereditas* 139:28–34

- Pons J, Bruvo B, Petitpierre E, Plohl M, Ugarković Đ, Juan C (2004) Complex structural features of satellite DNA sequences in the genus *Pimelia* (Coleoptera: Tenebrionidae): random differential amplification from a common "satellite DNA library." *Heredity* 92:418–427
- Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* 14:817–818
- Rojas-Rousse D, Bigot Y, Periquet G (1993) DNA insertions as a component of the evolution of unique satellite DNA families in two genera of parasitoid wasps: *Diadriomus* and *Eupelmus* (Hymenoptera). *Mol Biol Evol* 10:383–396
- Rozas J, Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15:174–175
- Schueler MG, Higgins AW, Rudd MK, Gustashaw K, Willard HF (2001) Genomic and genetic definition of a functional human centromere. *Science* 294:109–115
- Stephan W, Cho S (1994) Possible role of natural selection in the formation of tandem-repetitive noncoding DNA. *Genetics* 136:333–341
- Strachan T, Webb D, Dover GA (1985) Transition stages of molecular drive in multiple-copy DNA families in *Drosophila*. *EMBO J* 4:1701–1708
- Sun X, Le HD, Wahlstrom JM, Karpen GH (2003) Sequence analysis of a functional *Drosophila* centromere. *Genome Res* 13:182–194
- Swofford DL (1998) PAUP\*: Phylogenetic analysis using parsimony (\*and other methods), version 4. Sinauer Associates, Sunderland, MA
- Talbert PB, Masuelli R, Tyagi AP, Comai L, Henikoff S (2002) Centromeric localization and adaptive evolution of an *Arabidopsis* histone H3 variant. *Plant Cell* 14:1053–1066
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 24:4876–4882
- Ugarković Đ, Plohl M (2002) Variation in satellite DNA profiles — Causes and effects. *EMBO J* 21:5955–5959
- Ugarković Đ, Podnar M, Plohl M (1996a) Satellite DNA of the red flour beetle *Tribolium castaneum*—Comparative study of satellites from the genus *Tribolium*. *Mol Biol Evol* 13:1059–1066
- Ugarković Đ, Durajlija S, Plohl M (1996b) Evolution of *Tribolium madens* (Insecta, Coleoptera) satellite DNA through DNA inversion and insertion. *J Mol Evol* 42:350–358
- Volpe TA, Kidner C, Hall IM, Teng G, Grewal SIS, Martiensses RA (2002) Regulation of heterochromatic silencing and histone H3 Lysine-9 methylation by RNAi. *Science* 297:1833–1837
- Warburton PE, Willard HF (1990) Genomic analysis of sequence variation in tandemly repeated DNA. *J Mol Biol* 216:3–16
- Willard HF, Waye JS (1987) Hierarchical order in chromosome-specific human a-satellite DNA. *Trends Genet* 3:192–198