Parallelism in Evolution of Highly Repetitive DNAs in Sibling Species

Brankica Mravinac[1] and Miroslav Plohl*[,1]

[1]Division of Molecular Biology, Ruđer Bošković Institute, Zagreb, Croatia

***Corresponding author:** E-mail: plohl@irb.hr.

Associate editor: Patricia Beldae

Abstract

Characterization of heterochromatin in the flour beetle *Tribolium audax* revealed two highly repetitive DNA families, named TAUD1 and TAUD2, which together constitute almost 60% of the whole genome. Both families originated from a common ancestral ~110-bp repeating unit. Tandem arrangement of these elements in TAUD1 is typical for satellite DNAs, whereas TAUD2 represents a dispersed family based on 1412-bp complex higher-order repeats composed of inversely oriented ~110 bp units. Comparison with repetitive DNAs in the sibling species *Tribolium madens* showed similarities in nucleotide sequence and length of basic repeating units and also revealed structural and organizational parallelism in tandem and dispersed families assembled from these elements. In both *Tribolium* species, one tandem and one dispersed family build equivalent distribution patterns in the pericentromeric heterochromatin of all chromosomes including supernumeraries. Differences in the nucleotide sequence and in the complexity of higher-order structures between families of the same type suggest a scenario according to which rearranged variants of the corresponding ancestral families were formed and distributed in genomes during or after the speciation event, following the same principles independently in each descendant species. We assume that random effects of sequence dynamics should be constrained by organizational and structural features of repeating units and possible requirements for spatial distribution of particular sequence elements. An interspersed pattern of repetitive families also points to the intensive recombination events in heterochromatin. Synergy between the meiotic bouquet stage and satellite DNA sequence dynamics could make a positive feedback loop that promotes the observed genome-wide distribution. At the same time, considering the abundance of these DNAs in heterochromatin spanning the (peri)centromeric chromosomal segments, we speculate that diverged repetitive sequences might represent the DNA basis of reproductive barrier between the two sibling species.

Key words: satellite DNA, interspersed repeats, sibling species, *Tribolium*, heterochromatin, evolution.

Introduction

Tandemly repeated DNA sequences, known as satellite DNAs, constitute eukaryotic heterochromatin blocks, located mostly at centromeric and telomeric chromosomal positions (Charlesworth et al. 1994). These regions are poorly represented in outputs of genome sequencing projects as their repetitive nature complicates approachability to sequence assembly. Nowadays, it becomes evident that comprehension of genomes in their entirety depends on research of repetitive DNAs (Shapiro and von Sternberg 2005), and functional roles have been ascribed to sequences that were historically attributed as noncoding genomic excess (reviewed in Plohl et al. 2008; Malik and Henikoff 2009).

Sequences repeated in tandem fashion form arrays that sometimes span up to several megabases (Schueler and Sullivan 2006). These sequences evolve according to the principles of concerted evolution driven by different mechanisms of nonreciprocal DNA transfer, such as unequal crossover between repeating units, gene conversion, rolling circle replication followed by reinsertion, replicative transposition, and slippage replication (Dover 1986, 2002). Turnover mechanisms spreading and homogenizing concomitantly two or more adjacent repeats can generate higher-order repeats (HORs), characterized by higher sequence divergence between subunits within a HOR than among HORs themselves (Willard and Waye 1987; Warburton et al. 2008). At the interspecies level, evolutionary related organisms may share a common "library" of satellite sequences, whose quantitative changes can generate species-specific satellite composition (Meštrović et al. 1998 and references therein). Sequence rearrangements such as duplications, insertions, deletions, and inversions within monomers may play a dominant role in evolution of some satellite sequences (Mravinac and Plohl 2007), and rearranged variants follow the principles of the "library" model (Cesari et al. 2003; Pons et al. 2004).

The cosmopolitan genus *Tribolium* comprises 33 insect species, including the major global pests of stored grain and cereal commodities for human consumption. Potential of *Tribolium* as a powerful model system for genetic analyses and manipulations was recognized decades ago (Sokoloff 1972), having as an outcome the whole genome sequence data of the red flour beetle *Tribolium castaneum* (Richards et al. 2008). The genus itself consists of closely related and distantly related species (Sokoloff 1972), which provide an excellent platform for comparative and evolutionary studies. In terms of noncoding sequences, an extensive research has been done on highly repetitive DNA sequences in *Tribolium freemani* (Juan et al. 1993), *Tribolium confusum*

(Plohl et al. 1993), *T. castaneum* (Ugarković, Podnar, and Plohl 1996), *Tribolium anaphe* (Mravinac et al. 2004), *Tribolium destructor* (Mravinac et al. 2004), *Tribolium brevicornis* (Mravinac et al. 2005), and *Tribolium madens* (Ugarković, Durajlija, and Plohl 1996; Žinić et al. 2000; Mravinac and Plohl 2007). Single species–specific satellite DNAs dominate in the majority of *Tribolium* species making up to 40% of the whole genomes, whereas two major satellite DNA families were characterized in *T. madens*. The American flour beetle, *Tribolium audax*, and its sibling species *T. madens* are so similar in phenotype that for a long time they were classified as a single species. There are, however, slight morphological differences between them, and the only offspring they produce is limited to a few sterile hybrids (Halstead 1969). Given the two dominant satellite DNAs in *T. madens* and their distribution in pericentromeric regions of all chromosomes, the question of *T. audax* satellite DNA organization arises, especially in the light of potential role that these sequences could play in species divergence.

Here, we identified and analyzed the most abundant repetitive DNA families in the genome of *T. audax*. Comparison of *T. audax* repetitive DNA families and satellite DNAs of the sibling species *T. madens* undoubtedly points to their common origin. Based on correlations between repetitive DNAs in the two species, we conclude that the evolution of repetitive DNAs followed the same principles, promoting repeating units of comparable organizational and structural features and their distribution in pericentromeric heterochromatin of all chromosomes. This parallelism may be favored by similar requirements imposed on DNA sequences occupying equivalent genomic locations in sibling species.

Materials and Methods

Detection and Cloning of Repetitive Sequences

Tribolium audax insect culture was purchased from the Central Science Laboratory (Sand Hutton, York, UK). Total genomic DNA was isolated, digested with 32 restriction enzymes (*Acc*I, *Alu*I, *Bam*HI, *Bgl*II, *Bsm*I, *Bst*EII, *Cel*II, *Cla*I, *Dde*I, *Dra*I, *Eco*RI, *Eco*RV, *Hae*III, *Hin*dII, *Hin*dIII, *Hin*fI, *Hpa*II, *Mlu*I, *Nco*I, *Not*I, *Pst*I, *Pvu*I, *Pvu*II, *Rsa*I, *Sac*I, *Sal*I, *Sau*3AI, *Sfi*I, *Sma*I, *Taq*I, *Xba*I, and *Xho*I) (Roche), and electrophoresed on 1% agarose gel. The fragments from the most prominent *Hin*fI, *Cla*I, and *Rsa*I bands were ligated into the pUC18 vector and sequenced. Based on the sequenced restriction fragments, three outward-facing polymerase chain reaction (PCR) primers were constructed to extend the sequenced region of genomic DNA. PCR products were generated by using different combinations of the primers 1) 5′AATCACAATCTTACCACTTG3′, 2) 5′TTATTTCAAGCTTTTCTCGC3′, and 3) 5′CCCTAAATCGAGCCAAACTAA3′ (primer positions are indicated in supplementary fig. S1, Supplementary Material online). PCR amplifications included predenaturation at 94 °C for 2 min, 35 cycles of amplification (denaturation at 94 °C for 20 s, annealing at 55 °C for 15 s, and extension at 72 °C for 1 min), and final extension at 72 °C for 7 min. PCR amplicons were cloned using the pGEM-T Vector System I (Promega). Cloned restriction fragments and PCR products were sequenced by MWG-Biotech DNA Sequencing Service (Ebensberg, Germany). The sequences have been deposited in the GenBank database under the accession numbers GU220883–GU220905.

Sequence Analyses

Initial sequence editing and manipulations were done by using BioEdit v.7.0.5.3 (Hall 1999). Different modules from Lasergene software package v.7.0.0 (Dnastar) were employed in further analyses of repetitive sequences including pairwise and multiple alignments, contig assembly, dot plot analyses, and PCR primer design. Tandem repeats within the TAUD consensus sequence were located by using Tandem repeats finder v.4.00 (Benson 1999), and shorter direct and inverted repeats were identified using the online programs OligoRep (http://wwwmgs.bionet.nsc.ru/mgs/programs/oligorep/) and REPuter (http://bibiserv.techfak.uni-bielefeld.de/reputer/). DNA polymorphism and the distribution of variability were analyzed by using DnaSP v.4.10.9 (Rozas et al. 2003). Uncorrected *p*-distance matrices were calculated according to the multiple alignments assisted by ClustalX v.1.81 (Thompson et al. 1997). Phylogenetic analyses were performed by maximum parsimony method using PAUP* v.4.0b10 software (Swofford 2002). Heuristic searches, conducted under Tree Bisection-Reconnection branch swapping option with all characters weighted equally and gaps treated as missing data, were supported by bootstrap values computed from 1,000 iterations.

Southern Blot, Colony Lift, and Dot Blot Hybridization

Hybridization experiments were performed in 250 mM Na_2HPO_4 (pH 7.2)/1 mM ethylenediaminetetraacetic acid (EDTA)/20% sodium dodecyl sulfate (SDS)/0.5% blocking reagent with agitation overnight at different stringency temperatures from 60 °C to 68 °C (allowing 80% to >95% homology), whereas posthybridization washes were done in 20 mM Na_2HPO_4/1 mM EDTA/1% SDS at a temperature 2 °C below the hybridization temperature. Blot filters were probed with digoxigenin-labeled TAUD1 and TAUD2 sequences. TAUD1- and TAUD2-specific probes were prepared by random-primed labeling of TAUD fragments between 429–989 and 1805–2837 nucleotide positions, respectively (supplementary fig. S1, Supplementary Material online). Chemiluminescent detection was carried out using the alkaline phosphatase substrate CDP-Star (Roche). The amount of TAUD1 and TAUD2 sequences in *T. audax* genome was determined by dot blot analysis using a dilution series of genomic DNA and cloned TAUD1 and TAUD2 fragments, and the obtained signals were quantified using the Image J software (http://rsb.info.nih.gov/).

Pulsed Field Gel Electrophoresis

High molecular weight DNA was prepared and digested in agarose plugs according to Žinić et al. (2000). Pulsed field gel electrophoresis (PFGE) was carried out using a BioRad Chef-DR III apparatus. A total of 100-bp to 20-kb fragments were separated in 1.2% agarose, using 0.5× Tris-borate-EDTA

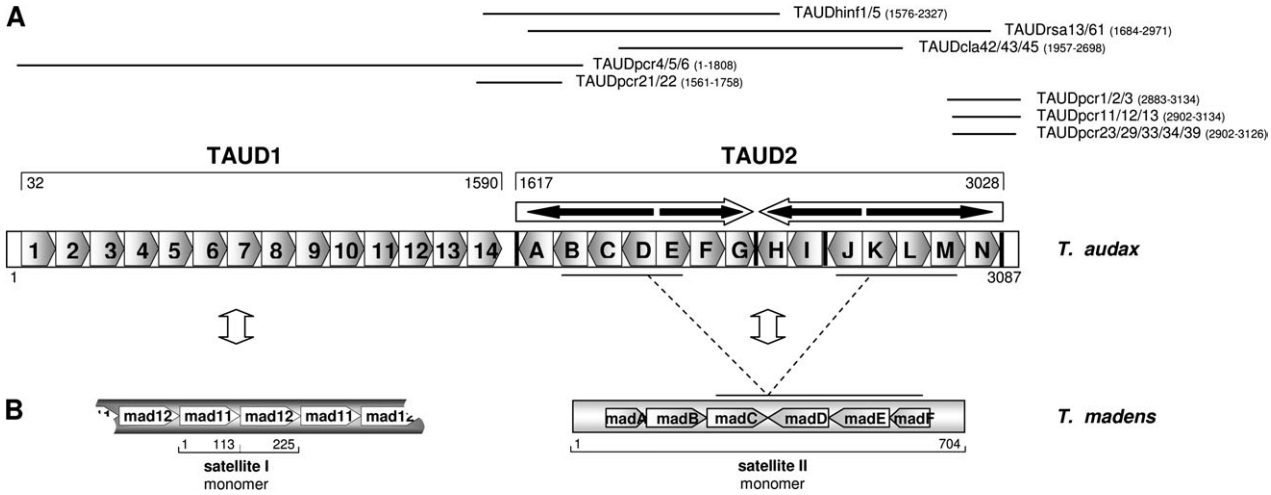

FIG. 1. (*A*) Schematic representation of structure and organization of the TAUD sequence. Cloned fragments, from which the TAUD sequence was derived, are indicated above the sequence scheme. Numbers in parentheses indicate fragment position within the TAUD consensus sequence shown in supplementary figure S1 (Supplementary Material online). Pentagons marked from 1 to 14 represent the TAUD1 monomers (nucleotide position 32–1590), whereas A–N pentagons represent the TAUD2 subunits (nucleotide positions 1617–3028). Each pentagon points to the monomer/subunit orientation. Inverted segments within TAUD2 are designated by arrows. Black bars within the TAUD2 scheme represent stretches that are not assigned to flanking subunits. (*B*) Schematic representation of *Tribolium madens* satellite I (225 bp) and satellite II (704 bp) monomer structure. Up–down arrows point to the orthologous families between *Tribolium audax* and *T. madens*. Dashed lines relate the most similar segments of TAUD2 and satellite II sequences.

(TBE) buffer, 0.1 s switch time, 120° reorientation angle, and 9 V/cm voltage at 14 °C during 3.5 h. Fragments in 0.1–2.2 Mb size range were separated on 1% agarose gels, using 0.5× TBE buffer, 60–120 s switch time, 120° reorientation angle, and 6 V/cm voltage at 14 °C during 25 h. Southern blot transfer and hybridization of PFGE fragments were done as described in the previous paragraph.

Two-Color Fluorescence In Situ Hybridization

Chromosome spreads were prepared from male gonads as described (Juan et al. 1991). TAUD1 and TAUD2 probes were labeled by nick translation using digoxigenin-11-2′-deoxy-uridine-5′-triphosphate (DIG-11-dUTP) and biotin-16-2′-deoxy-uridine-5′-triphosphate (biotin-16-dUTP), respectively. Hybridization was performed for 16 h at 37 °C in the solution containing 60% formamide, 2× saline-sodium citrate (SSC), 10% dextran sulfate, 20 mM sodium phosphate, and 5 ng/μl of each probe. High stringency posthybridization washes were done in 50% formamide/2× SSC at 42 °C. Immunological detection was done as detailed in Žinić et al. (2000). All signals were viewed through appropriate filters for blue (4′,6-diamidino-2-phenylindole [DAPI]), red (tetramethylrhodamine isothiocyanate), and green (fluorescein isothiocyanate) fluorescence, using Opton Leitz microscope equipped with Pixera Pro150ES digital camera. Images were acquired using Twain Viewfinder v.3.0 capture program.

Results

Detection and Characterization of TAUD1 and TAUD2 Repetitive Families

To detect repetitive DNA sequences in the genome of *T. audax*, we digested genomic DNA with a series of restriction enzymes. Cloned and sequenced fragments, extracted from prominent 750-bp *Hin*fI and *Cla*I bands, as well as from a 1300-bp *Rsa*I band, assembled a 1396-bp long sequence (fig. 1A). To extend the sequenced segment, specific PCR primers that match lateral positions on both sides of the segment were constructed. By combining different primers we amplified five groups of fragments that extended the initial 1396-bp sequence up to 3087 bp (fig. 1A). An alignment of the TAUD consensus sequence and the source fragments is listed in supplementary figure S1 (Supplementary Material online). Sequence analysis of the TAUD sequence disclosed juxtaposition of two structurally distinct repetitive families, denoted as TAUD1 and TAUD2, both based on ~110-bp basic repeating units (fig. 1A).

TAUD1 and TAUD2 Families

The first 1559 bp of the studied fragment (position 32–1590; fig. 1A) consists of 14 tandemly head-to-tail repeated units (aud1–aud14; fig. 1A), ranging in size from 104 to 113 bp (fig. 2), with an average nucleotide diversity (Pi) of 23.4%. This repetitive family is named TAUD1. Southern hybridization with a specific probe prepared from the TAUD1 array revealed the satellite DNA typical ladder-like profile based on a ~110-bp fragment and its multimers (fig. 3A).

Phylogenetic analysis classified TAUD1 monomers into two distinct clades. Based on 18 diagnostic nucleotides that they have in common (fig. 2), monomers aud2, aud7, and aud13 form a separate statistically well-supported group, whereas all other TAUD1 monomers fall into another clade (fig. 4). Hundred percent identity between 264-bp sequence fragments at nucleotide positions 168–431 and 729–992 implies traces of sequence duplication and/or putative gene conversion events. Restrictions of genomic DNA followed by Southern blot analysis (fig. 3A) did

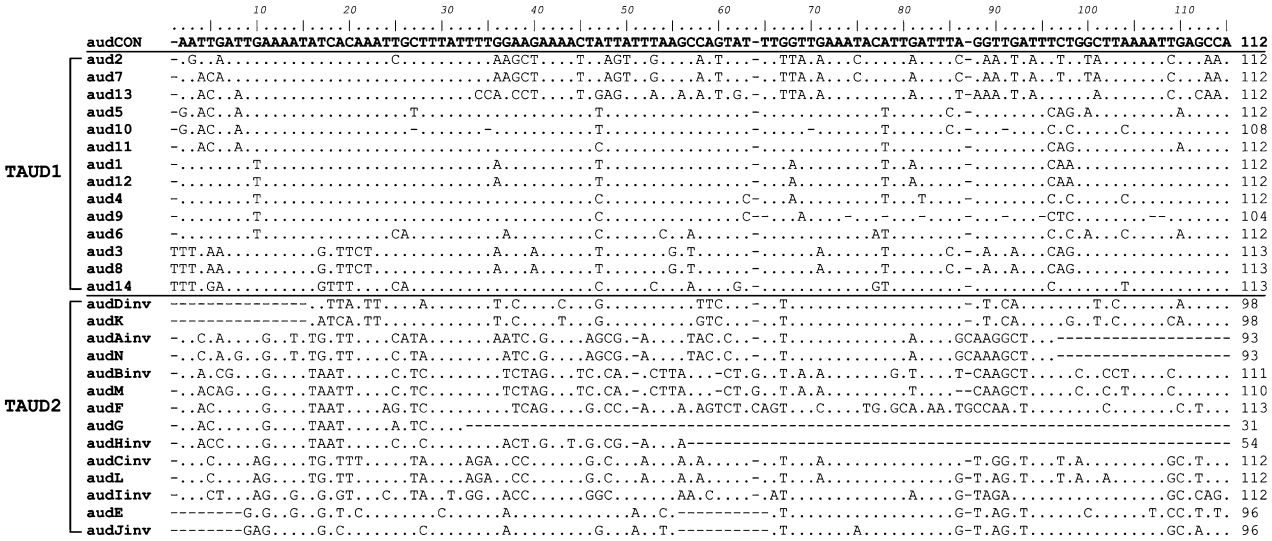

Fig. 2. Alignment of TAUD1 (aud1–aud14) and TAUD2 (audA–audN) repetitive elements. Reverse complementary notation of the TAUD2 subunits relative to their original orientation is indicted by "inv" suffix. The consensus sequence (audCON) representing the most prevailing nucleotide at each position is shown in bold; positions identical to the consensus are indicated by dots, whereas deletions are marked by dashed lines.

not indicate any HOR form that would constitute the prevailing organizational pattern of TAUD1 family, which therefore can be defined by ~112-bp-long monomer repeats. According to dot blot analysis (data not shown), TAUD1 makes 40.0% of *T. audax* genome. By means of *T. audax* C-value (Alvarez-Fuster et al. 1991), copy number is determined to be about 5.5×10^5 monomers per haploid genome.

A 1412-bp-long segment (position 1617–3028; fig. 1A) was designated as TAUD2. From organizational point of view, the TAUD2 sequence can be assorted into 14 subunits (audA–audN), whose schematic organization is given in fig. 1A. The subunit orientation delineates the two 698-bp-long inverted repeats (fig. 1A), whose 86.5% identity in nucleotide sequence could potentially support a hairpin structure formation. The subunit alignment points to their common origin (fig. 2), despite considerable variation in size (31–113 bp) and wide range of pairwise similarities (47.3–96.4%) (supplementary table S1, Supplementary Material online). In accordance with TAUD2 internal symmetry, the mirrored subunits (audA–audN, audB–audM, audC–audL, audD–audK, and audE–audJ) share the most similar sequences (fig. 4). The central subunits audF/G/H/I, however, reflect disrupted symmetry what makes one speculate that in a hypothetical hairpin structure requirements to preserve DNA sequence in the loop segment are weaker in comparison with the stem segments. Orientation of subunits (fig. 1A), together with their sequence similarities (fig. 2), indicates several inversion/duplication events before the TAUD2 element was formed. Fourteen subunits span 94.1% of the TAUD2 sequence, whereas the rest pertains to four stretches of related sequence segments (fig. 1A and supplementary fig. S1, Supplementary Material online). The boundaries of the TAUD2 sequence, that flank audA–audN subunits, are built by 24-bp-long segments, 95.8% identical in inverted orientation, showing up to 69.6% similarity to TAUD2 subunits. The two short stretches detected between audG–audH (15 bp) and audI–audJ (20 bp) share up to 80.0% similarity to particular segments within TAUD2 subunits.

Southern hybridization experiments using TAUD2-specific probe revealed distinct bands within digested genomic DNA (fig. 3B). However, TAUD2 complexity disabled us to find a restriction endonuclease that would define the whole HOR monomer and indicate its organization into tandem arrays (fig. 3B). Bands below 1.5 kb correspond to the fragments predicted by restriction analysis in silico, whereas higher bands probably reflect mutations within the TAUD2 sequence and flanking regions. We reason that such distribution mirrors predominantly interspersed organization of highly repetitive TAUD2 repeats. Using dot blot assay, we established that TAUD2 comprises 19.3% of *T. audax* genome (data not shown), corresponding to $\sim 2.1 \times 10^4$ copies per haploid complement.

Relations between TAUD1 and TAUD2 Repetitive Families

Sequence comparison (fig. 2) revealed that TAUD1 and TAUD2 110-bp basic repeating units distribute across two major clades in a phylogenetic tree (fig. 4). Interestingly, the clades do not reflect a strict classification of the families, and the two TAUD2 subunits, mirror-symmetrical audD and audK, reside within the TAUD1 clade. Relationship between building elements of the two families supports their common origin, indicating that audD/K pair might be the most approximate to the sequence of ancestral repetitive elements, from which both extant tandem repeats and complex inverted structures evolved.

Junction between TAUD1 and TAUD2 is defined by a 26-bp segment (fig.1A, positions 1591–1616) that

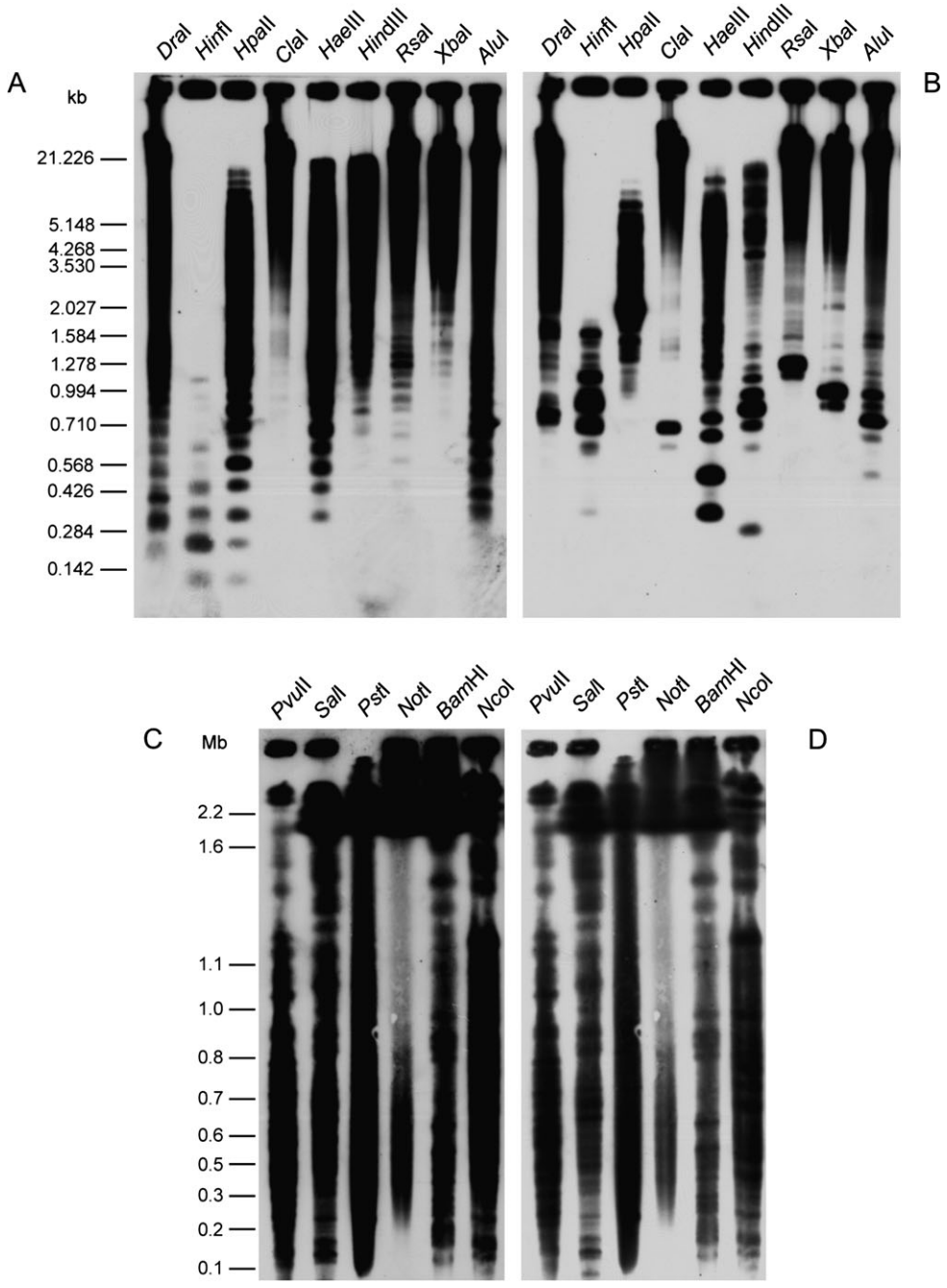

FIG. 3. Southern blots obtained after hybridization with the probe specific for TAUD1 (*A* and *C*) and for TAUD2 (*B* and *D*). Genomic DNA was digested by restriction enzymes indicated above the lanes. Fragments are separated in the range from 100 bp to 20 kb (*A* and *B*) and in the range from 0.1 to 2.2 Mb (*C* and *D*).

represents a truncated variant of ~110-bp repeating unit, being 48.1–92.3% similar to TAUD1 and TAUD2 repeats. Existing of alternative sequence arrangements, slightly different than the former one, is shown in clones TAUDpcr21 and TAUDpcr22 (supplementary fig. S1, Supplementary Material online). TAUD2 3′ flanking region (positions 3029–3087) also shares similarity with both TAUD1 (54.2–64.9%) and TAUD2 (60.4–87.5%) repeat elements. Beyond that point, the two alternative extensions (42 and 34 bp) indicate at least two different contexts that flank TAUD2 (TAUDa and TAUDb sequences, respectively, in supplementary fig. S1, Supplementary Material online). In terms of sequence origin, both extensions show partial similarity to TAUD1/TAUD2 repeat variants.

Genomic Distribution of TAUD1 and TAUD2 Families

Chromosomal localization of the two repetitive families was assessed by fluorescence in situ hybridization (FISH) to prophase and metaphase spreads. *Tribolium audax* karyotype comprises 9 autosomal pairs, an Xy_p pair of sex chromosomes associated in a parachute formation, and 3–18 supernumerary B-chromosomes (Shimeld 1989). DAPI staining confirmed the presence of large AT-rich

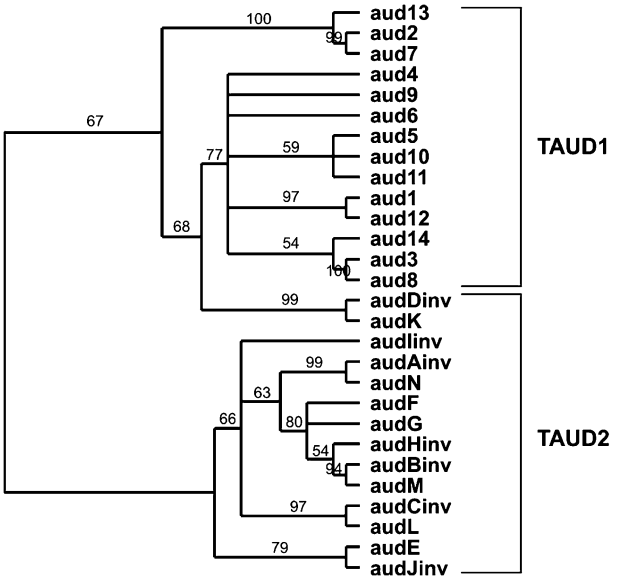

Fig. 4. Parsimony cladogram reflecting relationships between TAUD1 (aud1–aud14) and TAUD2 (audA–audN) repetitive elements. The suffix "inv" signifies inverse orientation of the TAUD2 subunits with respect to the sequenced segment. Numbers above the branches indicate percentages of bootstrap values for 1,000 replicates (values under 50% are not shown).

heterochromatic blocks in (peri)centromeric regions of the whole chromosome set, including supernumeraries (fig. 5*A*). Simultaneous two-color FISH assay with probes for TAUD1 (fig. 5*B*) and TAUD2 (fig. 5*C*) visualized uniform distribution of both families within heterochromatic domains of all chromosomes. In addition, in meiotic prophase I we noted bouquet-like arrangements of nonhomologous chromosomes where TAUD1 and TAUD2 sequences undoubtedly participate in centromeric associations (pointed by arrows in fig. 5).

Large-scale genomic distribution of TAUD1 and TAUD2 repeats was studied by PFGE followed by Southern blot hybridization (fig. 3*C* and *D* and supplementary fig. S2, Supplementary Material online). Using restriction enzymes that do not find restriction sites within examined sequences, we obtained fragments of *T. audax* genomic DNA in the separation range from 50 kb up to 2 Mb. Hybridization with TAUD1/TAUD2 probes revealed the most intensive signals in the region between 200 and 800 kb (fig. 3*C* and *D*). As there is no observed difference between TAUD1 and TAUD2 hybridization patterns, these fragments apparently involve elements of both repetitive families arranged in a combinatorial fashion. Comparison of hybridization profiles obtained after cleaving high–molecular weight DNA with restriction enzymes that cut either TAUD1 or TAUD2 sequences, or both of them (fig. 3*A* and *B*) confirmed almost completely overlapping hybridization profiles for fragments larger than 10 kb. The exception was observed in *Hpa*II digestion, which revealed ~15-kb-long arrays that contain TAUD1 monomers but are deprived of TAUD2. PFGE blots are in tune with FISH analysis, confirming the distribution of two families in a highly interspersed pattern within the same genomic regions. Notwithstanding TAUD1/TAUD2 juxtaposition within relatively short segments, we cannot exclude the possibility that the mosaic arrays of the two families include intervening extraneous DNA sequences.

Comparison of *T. audax* and *T. madens* Repetitive Families

Search for sequences similar to TAUD1 and TAUD2 in the nucleotide collection database of the National Center for

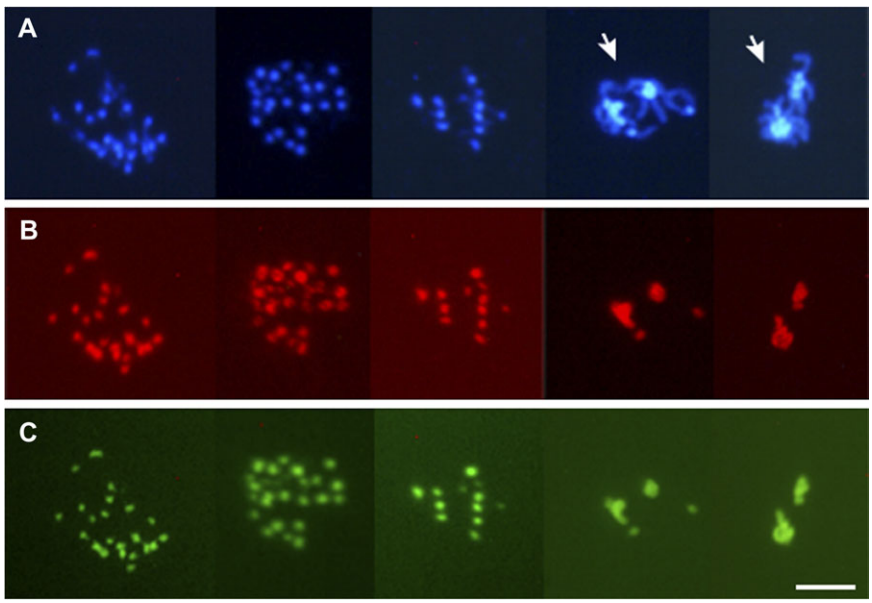

Fig. 5. FISH on *Tribolium audax* chromosomes ($2n = 18 + Xy_p +$ ss). Chromosomes stained with 4′,6-diamidino-2-phenylindole (*A*), hybridized with TAUD1 probe visualized by tetramethylrhodamine isothiocyanate (red fluorescence) (*B*), and hybridized with TAUD2 probe visualized by fluorescein isothiocyanate (green fluorescence) (*C*). The arrows point to chromosomes in meiotic bouquet stage. The bar represents 5 μm.

Biotechnology Information identified significant similarities only with satellite DNAs from the congeneric species *T. madens* (Ugarković, Durajlija, and Plohl 1996; Mravinac and Plohl 2007). TAUD1 consensus sequence recognizes the subunit mad11 of dimeric-based *T. madens* satellite I (fig. 1*B*) as the most homologous sequence (90.1% identity). The best basic local alignment search tool score of TAUD2 consensus sequence is against *T. madens* satellite II, a complex 704-bp sequence based on six subunits arranged in a mirror-symmetrical form (fig. 1*B*). A 352-bp segment that shows the most convincing sequence similarity (84.5%) between TAUD2 and *T. madens* satellite II includes four successive subunits, audB–audE/audJ–audM in TAUD2 and madC–madF in *T. madens* satellite II (fig. 1*B*). It is interesting to note that this segment traverses inversion points in which subunits change their orientation, both in TAUD2 (audD–audE and audJ–audK contacts) and in *T. madens* satellite II (madC–madD contact).

To gain detailed understanding of relationships between *T. audax* and *T. madens* repetitive DNAs, we compared all detected ~110-bp building elements of TAUD1 and TAUD2 (this work) with those of *T. madens* satellite I (Ugarković, Durajlija, and Plohl 1996) and satellite II (Mravinac and Plohl 2007). The sequences were aligned (supplementary fig. S3, Supplementary Material online), and the corresponding maximum parsimony cladogram is shown in figure 6. As expected, *T. madens* satellite I subunits mad11 and mad12 split into two groups. Intriguingly, *T. audax* aud2/7/13 monomers, which formed a separate branch in the TAUD1 clade (fig. 4), join mad12 subunits, whereas the rest of TAUD1 monomers group with mad11 subunits. This observation suggests that the two types of a ~110-bp repeating unit may have been differentiated before the *audax*/*madens* split, although they are represented in different abundance and organizational patterns in the two descendant satellites. On the other hand, variants of TAUD2 and *T. madens* satellite II subunits mutually intermingle (fig. 6). Statistically the best supported interspecies branch includes madD/audD/audK (fig. 6), the subunits adjacent to inversion points (fig. 1). Because the orthologous inversion points can be recognized in both species, we hypothesize that this inversion is ancestral and that in TAUD2 genesis it preceded an additional cycle of duplication/inversion event that formed ~700-bp-long inversely oriented segments. The fact that madD/audD/audK branch groups with mad11/TAUD1 sequences may indicate that madD/audD/audK subunits experienced the minimum number of changes during the formation of the ancestral complex unit.

To test if *T. audax* repetitive DNA sequences exist in their present form in *T. madens* and vice versa, we did the cross-hybridization experiments, in which *T. audax* and *T. madens* genomic DNAs were digested with the same restriction endonucleases and successively hybridized with four different probes under high and low stringency conditions (supplementary fig. S4, Supplementary Material online). Hybridization with TAUD1 probe, besides a regular TAUD1 ladder in *T. audax*, under low stringency conditions revealed only faint signals in *T. madens* DNA digestions, corresponding to longer fragments of *T. madens* satellite I family. A comparable result was yielded with *T. madens* satellite I probe. In *T. madens* digestion lanes, TAUD2 probe hybridized weakly just with the most intensive bands belonging to satellite II and, reversely, *T. madens* satellite II probe recognized slightly a very few TAUD2 fragments. Consequently, hybridization experiments did not reveal conserved forms of current repetitive families in the siblings.

Discussion

In this work we characterized TAUD1 and TAUD2, the two highly repetitive DNA families in *T. audax* that together make up about 60% of the genome. Their comparison with the abundant satellite I and satellite II in the sibling species *T. madens* (Ugarković, Durajlija, and Plohl 1996; Žinić et al. 2000; Mravinac and Plohl 2007) disclosed structural and organizational parallelism in DNA sequences of (peri)centromeric heterochromatin. Analogies can be identified at several levels. First, all four repetitive DNA families are founded on ~110-bp basic repeating units, clearly derived from a common ancestral sequence. Second, the four repetitive families are differentiated into two types according to their organization and sequence properties, where a single species comprises one family of each type. The first type consists of TAUD1 and satellite I, characterized by tandem arrangement of repeating units and their dominance in the genomes. The less abundant families of the second type, TAUD2 and satellite II, are based on complex HORs made of inversely oriented ~110-bp basic repeating units. Third, TAUD2 and satellite II copies are interspersed throughout TAUD1 and satellite I arrays, respectively, building an irregular pattern that stretches along the pericentromeric heterochromatin of all chromosomes in both species. DNA fragments composed exclusively of one sequence type are therefore relatively short, not longer than 15 kb in *T. audax* (this work) and up to 70 kb in *T. madens* (Žinić et al. 2000), resulting in numerous junction regions. These facts suggest that the ancestor of *T. audax* and *T. madens* comprised both types of repetitive DNAs, what implies persistence of the organizational pattern, although the nucleotide sequence in basic units differs remarkably. Nevertheless, the most intriguing evolutionary question remains how and why the irregular patchwork of the two types of repetitive elements persists in parallel in the two sibling species.

The TAUD1 satellite consists of ~110-bp monomers, although these elements are tandemly arranged as dimeric HORs of 225 bp in *T. madens* satellite I (Ugarković, Durajlija, and Plohl 1996). Phylogenetic relationship between TAUD1 monomers and satellite I subunits indicates origin of satellite I dimeric structure through HOR amplification of adjacent ~110-bp repeat variants present in an ancestral array. Formation of a HOR can be accompanied by inversion of whole subunit(s), as discussed in details for the 1061-bp satellite monomer of *T. brevicornis* (Mravinac et al. 2005 and references therein). TAUD2 and satellite

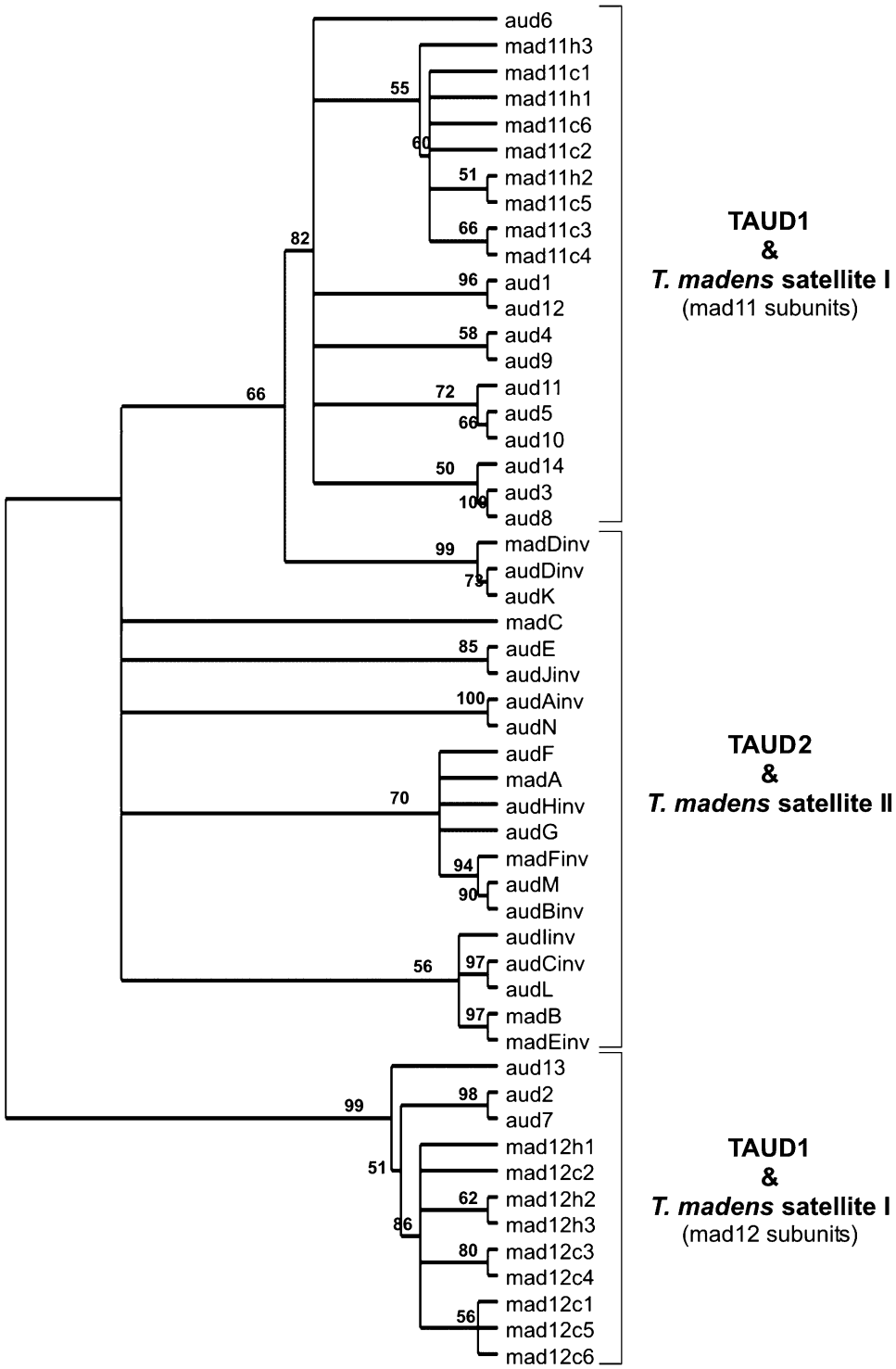

Fig. 6. Maximum parsimony tree of TAUD1 monomers (aud1–aud14) and TAUD2 subunits (audA–audN) cloned from *Tribolium audax* (this work), *Tribolium madens* satellite I subunits derived from individual monomers (mad11[c1–h3] and mad12[c1–h3]; Ugarković, Durajlija, and Plohl 1996), and *T. madens* satellite II subunits (madA–madF; Mravinac and Plohl 2007). The suffix "inv" marks reverse complementary orientation of subunits regarding the sequenced segments. Bootstrap values indicate percentage of branches presenting the particular topography after 1,000 bootstrap replicates (values under 50% are omitted).

II sequences may have evolved as rearranged and more complex variants of an ancient inverted repeat that existed in the ancestral species. Similarity between "switch-point" subunits in inverted repeats of both repetitive families (fig. 1*B*) speak in favor of this hypothesis.

Parallelism in *T. audax* and *T. madens* repetitive sequences, in particular parallelism in inverted repeat structure and its distribution as a less abundant interspersed variant, indicates that some features can be evolutionary favored. Several reports suggest putative interactions promoted by

inverted repeat–structured satellite DNAs capable to form potentially stable hairpin structures (Lorite et al. 2004; Jonstrup et al. 2008). Although inverted repeat structure may provide a functional signal, its interspersion may be needed to distribute putative signal elements uniformly along the heterochromatin. In the human α-satellite, studies of sequential distribution indicate that HOR arrangement is needed for proper spacing of functional centromere protein B–binding sites (Ikeno et al. 1994; Hall et al. 2003), suggesting that amplification in a higher-order register and homogenization of HOR structures in an array can be under selection.

In the majority of species analyzed so far, satellite families form separate arrays arranged in a sequential order, making chromosome-specific combinations, as in the case of *Drosophila* (Lohe et al. 1993) and primates (Alexandrov et al. 2001). In beetles, satellite DNAs distributed on all chromosomes including supernumeraries (Petitpierre et al. 1995; Žinić et al. 2000; this work) point to the specificities in heterochromatin sequence dynamics and, consequently, organization. In addition, random distribution of satellite monomer variants in pericentromeric regions of all chromosomes, as noted in *Tenebrio molitor* (Plohl et al. 1992; Bruvo et al. 1995), could be a general feature of satellite DNAs in beetles. As such, it seems that interchromosomal barrier to the homogenization process is not a limiting parameter in the evolution of tenebrionid satellites studied so far. In the case of two (or more) satellite families, the same principle would lead to random interspersion of satellite arrays, as described above for *T. audax* and *T. madens*. As documented here, an average array size of a homogeneous sequence reflects the relative abundance of each constituent repeat in the genome: More abundant sequence makes longer homogeneous arrays of tandem repeats.

It is tempting to speculate that high abundance and similarity among building elements of *audax*/*madens* repetitive families may be provocative enough for recombination machinery to induce intensive rearrangements in the heterochromatic portion of the genome. It has been evidenced that recombination between heterologous repeats of human α-satellite tolerates up to 30% sequence divergence (Okumura et al. 1987), which is a criterion that most of *audax*/*madens* repetitive (sub)units meet. Illegitimate recombination and unequal crossover were suggested as mechanisms that established highly interspersed pattern of two unrelated satellites localized on microchromosomes of three *Drosophila* species (Kuhn et al. 2009). Junctions detected in the study of *T. madens* satellites are also direct, consistent with similar predicted "breakage-and-rejoining" pathway, although the effect of unequal crossover on sequence variability was observed in one junction type (Mravinac and Plohl 2007). In *T. audax* (this work), the transition region noted in the sequenced fragment comprises a short intervening sequence composed of degenerated fragments of the basic 110-bp repeating unit. Although this transition region may be unique, close association between the two families is additionally implied by PFGE experiments. Mechanisms other than illegitimate recombination and unequal crossover may contribute as well. Excision, rolling circle replication, and reinsertion of extrachromosomal circular DNA (eccDNA) can participate in the turnover of tandem repeats (Navratilova et al. 2008; Cohen and Segal 2009). Irregular patchwork of the two satellites can be established by multiple cycles of eccDNA turnover, resulting in segmental duplications and in different HORs. Transposon-mediated exchange is another mechanism likely to promote tandem repeat formation and dynamics in different organisms (Palomeque et al. 2006; Macas et al. 2009). It may be sufficient that inverted repeats in a satellite monomer resemble the structure of miniature inverted-repeat transposable elements to be recognized by enzymes responsible for transposition (Pons 2004). Whatever would be the mechanism, enhanced efficiency in the sequence spread between nonhomologous chromosomes may be facilitated by a bouquet formation occurring during the first meiotic division, as proposed previously (Žinić et al. 2000). This stage was observed in all studied tenebrionid species, belonging to the several genera, and it is probably a characteristic of the family (Žinić et al. 2000; fig. 5 in this work; Mravinac B, Plohl M, unpublished results). The synergy can be proposed to exist between the bouquet stage and satellite DNA sequence dynamics. In other words, sequence similarity in pericentromeric heterochromatin is required for alignments of nonhomologous chromosomes, although close associations among chromosomes are needed for highly efficient genome-wide dispersal of satellite DNA sequences.

In the context of speciation, genetic differences present one of the key factors of reproductive isolation (Coyne and Orr 1998). Henikoff and collaborators (Henikoff et al. 2001; Malik and Henikoff 2009) argue that evolution of centromeric repetitive DNAs results ultimately in differentiation of new species. It has recently been shown that heterochromatic repetitive DNA maintains the reproductive isolation of *Drosophila melanogaster* from its sibling, *Drosophila simulans* (Ferree and Barbash 2009). Diverged about 7 My ago (Meštrović et al. 2006), the sibling species *T. audax* and *T. madens* still can mate. However, they bring forth infertile F_1 hybrids (Halstead 1969), whose inviability and sterility point to mechanisms of postzygotic isolation. Concerning the portion of the genome that here described repetitive DNA families constitute and their dominance in heterochromatic regions spanning the (peri)centromere, it could be hypothesized that an acquisition of changes in these sequences could obstruct proper chromosome segregation, setting thus genetically reproductive barriers that finally lead to the speciation.

In conclusion, the phenomenon of equivalent organizational patterns in heterochromatin of the two *Tribolium* sibling species mirrors a complex network of interactions that are in the same time the cause and the consequence of the observed sequence distribution. Organizational and structural properties of satellite monomers, possible functional constraints imposed on sequence elements, copy number to which each family was amplified, and general

features of sequence dynamics in both genomes can be taken as major parameters defining heterochromatin architecture on the DNA level. The whole pericentromeric heterochromatin in *T. audax*, *T. madens*, and probably in all tenebrionid beetles as well acts as a uniform homogenization domain on all chromosomes irrespectively of the satellite DNA family. Results presented here favor the idea about high dynamics of recombination events in heterochromatin of at least some species/chromosome types, which contrasts the long-time established view on heterochromatin as recombinationally inert genomic compartment.

Supplementary Material

Supplementary figures S1–S4 and table S1 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by Research Fund of Republic of Croatia, project no. 098-0982913-2756 to M.P. We thank Barbara Mantovani and Nevenka Meštrović for valuable comments on the manuscript.

References

Alexandrov I, Kazakov A, Tumeneva I, Shepelev V, Yurov Y. 2001. Alpha-satellite DNA of primates: old and new families. *Chromosoma* 110:253–266.

Alvarez-Fuster A, Juan C, Petitpierre E. 1991. Genome size in *Tribolium* flour-beetles: inter- and intraspecific variation. *Genet Res.* 58:1–5.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27:573–580.

Bruvo B, Plohl M, Ugarković Đ. 1995. Uniform distribution of satellite DNA variants on the chromosomes of tenebrionid species *Alphitobius diaperinus* and *Tenebrio molitor*. *Hereditas* 123:69–75.

Cesari M, Luchetti A, Passamonti M, Scali V, Mantovani B. 2003. Polymerase chain reaction amplification of the Bag320 satellite family reveals the ancestral library and past gene conversion events in *Bacillus rossius* (Insecta, Phasmatodea). *Gene* 312:289–295.

Charlesworth B, Sniegowski P, Stephan W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371: 215–220.

Cohen S, Segal D. 2009. Extrachromosomal circular DNA in eukaryotes: possible involvement in the plasticity of tandem repeats. *Cytogenet Genome Res.* 124:327–338.

Coyne JA, Orr HA. 1998. The evolutionary genetics of speciation. *Philos Trans R Soc Lond B Biol Sci.* 353:287–305.

Dover G. 2002. Molecular drive. *Trends Genet.* 18:587–589.

Dover GA. 1986. Molecular drive in multigene families—how biological novelties arise, spread and are assimilated. *Trends Genet.* 2:159–165.

Ferree PM, Barbash DA. 2009. Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. *PLoS Biol.* 7:e1000234.

Hall SE, Kettler G, Preuss D. 2003. Centromere satellites from *Arabidopsis* populations: maintenance of conserved and variable domains. *Genome Res.* 13:195–205.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.* 41:95–98.

Halstead DGH. 1969. A new species of *Tribolium* from North America previously confused with *Tribolium madens* (Charp.) (Coleoptera: Tenebrionidae). *J Stored Prod Res.* 4:295–302.

Henikoff S, Ahmad K, Malik HS. 2001. The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* 293: 1098–1102.

Ikeno M, Masumoto H, 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.

Jonstrup AT, Thomsen T, Wang Y, Knudsen BR, Koch J, Andersen AH. 2008. Hairpin structures formed by alpha satellite DNA of human centromeres are cleaved by human topoisomerase IIα. *Nucleic Acids Res.* 36:6165–6174.

Juan C, Gosalvez J, Mezzanotte R, Petitpierre E. 1991. Cytological and biochemical characterization of the *in situ* endonuclease digestion of fixed *Tenebrio molitor* chromosomes. *Chromosoma* 100:432–438.

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.

Kuhn GC, Teo CH, Schwarzacher T, Heslop-Harrison JS. 2009. Evolutionary dynamics and sites of illegitimate recombination revealed in the interspersion and sequence junctions of two nonhomologous satellite DNAs in cactophilic *Drosophila* species. *Heredity* 102:453–464.

Lohe AR, Hilliker AJ, Roberts PA. 1993. Mapping simple repeated DNA sequences in heterochromatin of *Drosophila melanogaster*. *Genetics* 134:1149–1174.

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.

Macas J, Koblizkova A, Navratilova A, Neumann P. 2009. Hypervariable 3′ UTR region of plant LTR-retrotransposons as a source of novel satellite repeats. *Gene* 448:198–206.

Malik HS, Henikoff S. 2009. Major evolutionary transitions in centromere complexity. *Cell* 138:1067–1082.

Meštrović N, Mravinac B, Plohl M, Ugarković Đ, Bruvo-Mađarić B. 2006. Preliminary phylogeny of *Tribolium* beetles (Coleoptera: Tenebrionidae) resolved by combined analysis of mitochondrial genes. *Eur J Entomol.* 103:709–715.

Meštrović N, Plohl M, Mravinac B, Ugarković Đ. 1998. Evolution of satellite DNAs from the genus *Palorus*—experimental evidence for the "library" hypothesis. *Mol Biol Evol.* 15:1062–1068.

Mravinac B, Plohl M. 2007. Satellite DNA junctions identify the potential origin of new repetitive elements in the beetle *Tribolium madens*. *Gene* 394:45–52.

Mravinac B, Plohl M, Ugarković Đ. 2004. Conserved patterns in the evolution of *Tribolium* satellite DNAs. *Gene* 332:169–177.

Mravinac B, Ugarković Đ, Franjević D, Plohl M. 2005. Long inversely oriented subunits form a complex monomer of *Tribolium brevicornis* satellite DNA. *J Mol Evol.* 60:513–525.

Navratilova A, Koblizkova A, Macas J. 2008. Survey of extrachromosomal circular DNA derived from plant satellite repeats. *BMC Plant Biol.* 8:90.

Okumura K, Kiyama R, Oishi M. 1987. Sequence analyses of extrachromosomal Sau3A and related family DNA: analysis of recombination in the excision event. *Nucleic Acids Res.* 15:7477–7489.

Palomeque T, Antonio Carillo J, Muñoz-López M, Lorite P. 2006. Detection of a mariner-like element and a miniature inverted-repeat transposable element (MITE) associated with the heterochromatin from ants of the genus *Messor* and their

possible involvement for satellite DNA evolution. *Gene* 371:194–205.

Petitpierre E, Juan C, Pons J, Plohl M, Ugarković Đ. 1995. Satellite DNA and constitutive heterochromatin in tenebrionid beetles. In: Brandham PE, Bennett MD, editors. Kew chromosome conference IV. Kew (UK): Royal Botanic Gardens. p. 351–362.

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, Luchetti A, Meštrović N, Mantovani B. 2008. Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. *Gene* 409:72–82.

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.

Pons J. 2004. Cloning and characterization of a transposable-like repeat in the heterochromatin of the darkling beetle *Misolampus goudoti*. *Genome* 47:769–774.

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.

Richards S, Gibbs RA, Weinstock GM, et al. (238 co-authors). 2008. The genome of the model beetle and pest *Tribolium castaneum*. *Nature* 452:949–955.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497.

Schueler MG, Sullivan BA. 2006. Structural and functional dynamics of human centromeric chromatin. *Annu Rev Genomics Hum Genet*. 7:301–313.

Shapiro JA, von Sternberg R. 2005. Why repetitive DNA is essential to genome function. *Biol Rev Camb Philos Soc*. 80:227–250.

Shimeld L. 1989. A cytogenetic examination of eight species of *Tribolium*. *Tribolium Inf Bull*. 29:102–107.

Sokoloff A. 1972. The biology of *Tribolium* with special emphasis on genetic aspects. Oxford: Oxford University Press.

Swofford DL. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods). Sunderland (MA): Sinauer Associates.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTALX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. 25:4876–4882.

Ugarković Đ, Durajlija S, Plohl M. 1996. Evolution of *Tribolium madens* (Insecta, Coleoptera) satellite DNA through DNA inversion and insertion. *J Mol Evol*. 42:350–358.

Ugarković Đ, Podnar M, Plohl M. 1996. Satellite DNA of the red flour beetle *Tribolium castaneum*-comparative study of satellites from the genus *Tribolium*. *Mol Biol Evol*. 13:1059–1066.

Warburton PE, Hasson D, Guillem F, Lescale C, Jin X, Abrusan G. 2008. Analysis of the largest tandemly repeated DNA families in the human genome. *BMC Genomics*. 9:533.

Willard HF, Waye JS. 1987. Chromosome-specific subsets of human alpha satellite DNA: analysis of sequence divergence within and between chromosomal subsets and evidence for an ancestral pentameric repeat. *J Mol Evol*. 25:207–214.

Žinić SD, Ugarković Đ, Cornudella L, Plohl M. 2000. A novel interspersed type of organization of satellite DNAs in *Tribolium madens* heterochromatin. *Chromosome Res*. 8:201–212.