

# Structure and population dynamics of the major satellite DNA in the red flour beetle *Tribolium castaneum*

Isidoro Feliciello · Gianni Chinali · Đurđica Ugarković

Received: 5 May 2011 / Accepted: 1 August 2011 / Published online: 12 August 2011  
© Springer Science+Business Media B.V. 2011

**Abstract** In the beetle genus *Tribolium*, satellite DNAs comprise a significant amount of pericentromeric heterochromatin and are characterized by rapid turnover resulting in species specific profiles. In the present work we characterize the major pericentromeric satellite DNA TCAST of the beetle *T. castaneum* and analyse its population dynamics. Using direct sequencing of genomic PCR products we show that the TCAST satellite exists in the form of two related subfamilies: Tcast1a and Tcast1b that make up 20 and 15% of the genome, respectively. Tcast1a and Tcast1b have consensus sequences of 377 and 362 bp respectively, share an average similarity of 79% and are characterized by a divergent, subfamily specific region of approximately 100 bp. The two subfamilies are prevalently organized in the interspersed form, although a portion exists in the form of homogenous tandem arrays composed of only Tcast1a or Tcast1b. The pattern of restriction enzyme digestion indicates that Tcast1a and Tcast1b are organized in composite higher order repeats. Comparison of sequence variability of Tcast1a and Tcast1b among ten strains reveals a difference in the frequency of particular mutations present at some positions. However, no difference in the organization and in the amount of subfamilies was detected among strains. The results show that direct genomic sequencing can be a useful method for the detection of population specific features of satellite DNA.

In the case of TCAST satellite DNA, changes in the mutational profiles seem to represent the first step in the genesis of a population specific satellite profile.

**Keywords** Satellite DNA · Heterochromatin · Concerted evolution · *Tribolium castaneum*

## Introduction

Satellite DNAs are tandemly repeated sequences that are present as long uninterrupted arrays in heterochromatic regions of most eukaryotes (Charlesworth et al. 1994; Ugarković and Plohl 2002). In the insect genus *Tribolium* (Tenebrionidae, Coleoptera) cytogenetic studies of eight species revealed a conserved karyotype of 20 or 18 chromosomes characterized by the presence of large blocks of pericentromeric heterochromatin (Juan and Petitpierre 1991). Molecular analyses suggested that these blocks are composed almost exclusively of satellite DNAs that comprise up to 40% of the whole genome and encompass the regions of functional centromeres (Ugarković et al. 1996; Durajlija Žinić et al. 2000). *Tribolium* species have sequence specific satellite DNA profiles: in each species a single highly represented satellite is detected and the satellites exhibit no significant sequence similarity except common structural features in the form of stable dyad structures and A + T rich blocks (Plohl et al. 1993; Ugarković et al. 1996; Mravinac et al. 2004). Due to the high turnover of *Tribolium* satellite DNAs it may be expected that the divergence of their profiles precedes the speciation process and could be detected even at the population level.

In the red flour beetle *T. castaneum*, TCAST satellite DNA has been previously characterized as the major

I. Feliciello · Đ. Ugarković (✉)  
Department of Molecular Biology, Ruđer Bošković Institute,  
Bijenička 54, HR-10000 Zagreb, Croatia  
e-mail: ugarkov@irb.hr

I. Feliciello · G. Chinali  
Dipartimento di Medicina Clinica e Sperimentale,  
Università degli Studi di Napoli Federico II, via Pansini 5,  
I80131 Naples, Italy

satellite that comprises 17% of the whole genome (Ugarković et al. 1996). Since the haploid genome size of *T. castaneum* is 0.2 pg or 204 Mb (Alvarez-Fuster et al. 1991), this corresponds to around 80,000 copies (Ugarković et al. 1996; Wang et al. 2008). The satellite has a monomer size of 360 bp as revealed by sequencing of five randomly picked monomers, and encompasses centromeric as well as pericentromeric regions of all 18 chromosomes (Ugarković et al. 1996). Although the genome sequencing project of *T. castaneum* has been recently completed (Richards et al. 2008), most of the heterochromatic regions that are prevalently composed of TCAST satellite were excluded from the genome sequence due to technical difficulties associated with sequencing and assembling highly repetitive regions. Therefore, the number of clones in the GenBank sequence database that contain tandemly arranged TCAST repeats is very limited. Satellite DNA sequences characterized in different *Tribolium* species seem to be highly homogenous with the average variability of approximately 2% and absence of subfamilies, as revealed by comparison of randomly cloned monomers (Mravinac et al. 2004).

*Tribolium castaneum* is an important pest closely associated with stored cereal products or living in flour or feed mills. Although *T. castaneum* originated from India (Hinton 1948), it has been widely scattered by man, so that now it is considered cosmopolitan. For the study of *T. castaneum* satellite DNA sequence evolution at the population level, we used set of wild-type strains originating from diverse geographic locations in North- and South America, Europe and Japan (see “Materials and methods”). These locations are characterized by distinct environments ranging from continental and humid climate (Indiana, Missouri, USA), Mediterranean climate (Jerez, Spain) to subtropical and tropical climate (Kyushu, Japan, Peru, Ecuador). Samples of beetles were collected in a period from 1982 to 2010, and maintained as laboratory cultures.

In order to get more comprehensive insight into the structure, organization and variability of *T. castaneum* abundant satellite DNA we decided to sequence satellite DNA directly from genomic PCR products. Using this approach the average genomic sequence that represents the overall variability of the repeats is defined and the most variable positions in the satellite DNA are detected and characterized. The same method was previously used for the study of variability of satellite DNAs in frogs (Picariello et al. 2002; Feliciello et al. 2005) and rodents from the genus *Ctenomys* (Ellingsen et al. 2007; Caraballo et al. 2010). This method has some possible limitations such as primer specificity that can induce bias towards some variants while some others can remain hidden and

underrepresented. This limitation can be partially avoided by the use of relaxed annealing temperatures during PCR and multiple primers, some of them with degenerated positions (see “Materials and methods”). The pattern of overall variability of particular satellite DNAs or satellite subfamilies could reveal evolutionary processes affecting these sequences while comparison of variability among different strains and populations of the same species could be useful to follow satellite DNA dynamics from the population to species level. In addition to sequence variability, changes in the organization and abundance of particular satellites can be examined. By combining these approaches, it could be possible to reconstruct the scenario leading to the formation of species-specific heterochromatin, characteristic for *Tribolium* species.

Using direct genomic sequencing of TCAST satellite we showed that two related subfamilies Tcast1a and Tcast1b, together make up a significant portion of the genome and represent the major DNA component of heterochromatin of *T. castaneum*. The pattern of variability of the two subfamilies was characterized, as well as the organization and compartmentalization into more complex repeat units. Comparison among ten laboratory strains from different geographical locations revealed differences in mutational profiles, while no significant difference in the organization and copy number of the satellite subfamilies was detected.

## Materials and methods

### Beetle strains

The following strains of *T. castaneum* were used: GA-2 strain, originally used in the genome sequencing project and deriving from North American wild-type strain collected in Georgia in 1982, obtained from Dr. Dick Beeman (Manhattan, KS, USA); laboratory strain purchased from the Central Science Laboratory (CSL, Sand Hutton, York, UK) of European origin; seven wild-type strains obtained from Dr. Joachim Kurtz (Germany): 43—collected at Kyushu Island, Japan in 1988; 50—collected in Schegel Farm, Indiana, USA in 2005; 51—collected in Adrian, Missouri, USA in 2006; 52—collected in Bloomington, Indiana, USA in 2006; 55—collected in Jerez, Spain in 1991; 57—collected in Perù, in 2002; 61—collected in Banos, Ecuador in 1996; and a wild population collected near Zagreb, Croatia in 2010.

Each laboratory stock was established and maintained as a separate culture at a population size of >200 individuals on standard medium (20:1, flour:brewer’s yeast, by weight) in a dark incubator at 24°C and approximately 70% relative humidity.

## PCR amplification

Because of the tandem organization of satellite DNA, whole Tcast1a and Tcast1b repetitive units could be amplified by PCR using primers with the same origin in the repeat sequence, but with opposite orientation. Different sets of primers were used (see Fig. 1 for schematic positions):

pr1 (GGACCAACCATAAGCGAG),  
 pr2 (GGTCYTACAARAAAATGC),  
 pr3 (CGTCTCAAACATTGAAAAC TG),  
 pr4 (GACGAAGACGAAAATGCAGC),  
 a1 (GCAARCGCTTWATTAAGGTAACAG),  
 a2 (GCGYTTGCATATAAGCARATTTAG),  
 b1 (CATTTTTGTTCWGTAACWGTCACAG),  
 b2 (GTAATTCAGTTTTCAATGTATGAG),  
 a3 (CAGAATCGTTTTTAGTCGTC),  
 a4 (ACTGTTACCTTAATAAAGCG),  
 b3 (TAAATGCTGCAGTTTCGTC),  
 b4 (GTAAAACTGTTACCGTTTC),  
 a5 (CGAGATATAAGCAARAAACC),  
 b5 (TATAAGCAAAAACCAAAGTG).

The primer pairs indicated with “a” or “b” are specific for Tcast1a and Tcast1b respectively. The PCR reaction consisted of a classical 40 cycles, starting from 30 to 50 ng of genomic DNA with annealing temperature of 55 or 52°C, depending on primer pairs.

## Genomic sequencing

The average genomic sequence is obtained by direct amplification of genomic DNA using primer pairs corresponding to different regions of satellite monomer. The direct and inverse sequence electropherograms of each amplified repeat were compared and only bases present in both orientations were considered. The relative amount of each base in the positions containing two or more bases was considered to be proportional to the signal height. The average genomic sequences were converted into a genomic consensus sequence using IUBMB single letter code for

multiple bases in the same position (e.g. Y = C + T, K = T + G, W = A + T, S = G + C, R = A + G, M = A + C, D = A + T + G, H = A + T + C). A minor base was included in the genomic consensus sequence only if it gave a signal at least one-tenth that of the main base. The most common sequence (MCS) was obtained by indicating the predominant base found at each variable position. If there was no predominant base at a particular position, and if a minor base gave at least 40% of the principal signal height, IUBMB single letter code for multiple bases was used. The sequencing was performed using the automatic sequencer ABI Prism 310 (Applied Biosystems).

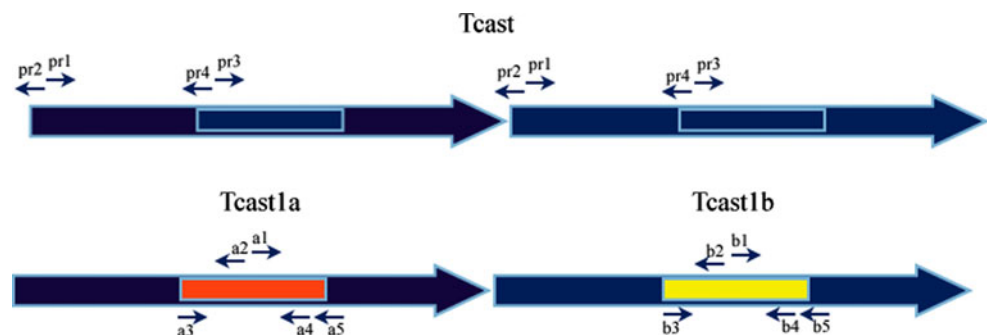
## Southern and dot blot hybridization

DNA was extracted from 50 to 100 mg of adult insects, which corresponds to 10–20 individuals, following the instruction of DNeasy Blood & Tissue Kit (QUIAGEN) for high molecular weight DNA purification. DNA was digested with different restriction enzymes (Fermentas-Life Science and SibEnzyme) under standard conditions, transferred to membrane and hybridized under high stringency conditions at 68°C.

For quantitative dot-blot analysis DNA samples at 1 µg/ml in denaturing buffer (0.4 M NaOH, 1 M NaCl) were serially diluted with an equal volume of herring sperm DNA (Promega) at the same concentration of 1 µg/ml in denaturing buffer. As a standard, a purified sample of a specific PCR-amplified Tcast1a or Tcast1b from genomic DNA was used at an initial concentration of 0.4 µg/ml + 0.6 µg/ml of herring sperm DNA in denaturing buffer. A 100 µl aliquot of each dilution was added per slot in a Dot-Blot apparatus (Bio-Rad), filtered on a nylon membrane Nytran N (Amersham Hybond—GE Healthcare) and fixed on a membrane by baking at 80°C. Hybridization and staining procedures were the same as those used for Southern blot analysis.

Biotinylated probes of the Tcast1a or Tcast1b satellite DNAs were obtained by PCR amplification of genomic DNA using primer pairs specific of each satellite DNA

**Fig. 1** Schematic representation of monomers of satellite DNA subfamilies Tcast1a (specific region is marked in red) and Tcast1b (specific region is marked in yellow) and positions of PCR primers used for determination of genomic sequence as well as for study of organization of satellite repeats





obtained by indicating the predominant base at each variable position (Fig. 2). Only at one position (237) in Tcast1a sequence bases A and G were present in almost same amount, and this position was indicated by IUBMB single letter code as “R” in the MCS. The results show that TCAST satellite DNA exists in the form of two subfamilies, Tcast1a and Tcast1b that have consensus sequences of 377 and 362 bp, respectively. Tcast1b corresponds to the previously characterized TCAST consensus sequence (Ugarković et al. 1996) while Tcast1a represents a new subfamily. Tcast1a and Tcast1b have mutual identity of 79%, but each has a sequence-specific region between positions 170 and 260 nt (Fig. 2).

#### Variability and organization of Tcast1a and Tcast1b subfamilies

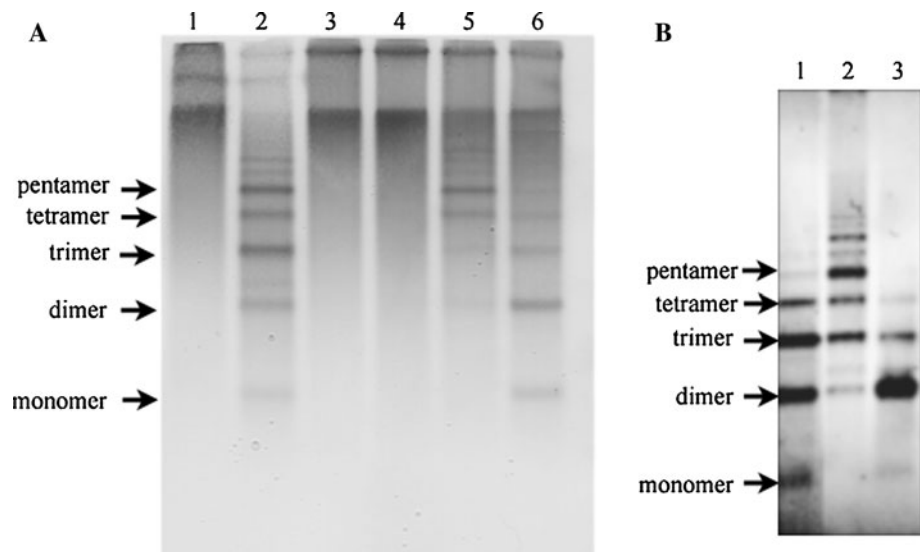
The organization and sequence variability of Tcast1a and Tcast1b satellite subfamilies were studied by Southern hybridization of DNA isolated from the GA-2 strain. Southern analysis represents a powerful method to validate the “genomic sequence” obtained by direct sequencing of a mixture of PCR-amplified repeats. Genomic DNA was digested with six restriction enzymes that recognize sequences of six base pairs: *AatII*, *Eco47III*, *NsiI*, *PaeI*, *PstI* and *ScaI*. Based on genomic consensus sequences and most common sequences of satellite subfamilies, *Eco47III* and *ScaI* have a single recognition site in Tcast1a, a single recognition site for *PstI* is present in a portion of Tcast1b repeats, while the recognition site for *AatII* does not occur within any of subfamilies (Fig. 2). Recognition sites for *NsiI* and *PaeI* can be created in Tcast1a by single point mutation. Blot was hybridized with mixed biotin labelled probes of Tcast1a and Tcast1b monomers under high stringency conditions (Fig. 3a). A hybridization ladder-like profile characteristic for tandemly repeated

satellite DNA is clearly visible on *T. castaneum* DNA digested with enzymes predicted to cut within Tcast1a genomic consensus sequence, while in the *PstI* digest, a faint ladder starting from the satellite dimer is visible (Fig. 3a). The restriction enzymes *AatII*, *NsiI* and *PaeI* do not produce a detectable ladder profile characteristic for satellite DNA. Digestion with restriction enzymes *ScaI* and *Eco47III* that cut once in Tcast1a gave a band of monomer size as well as a more prominent band of dimer size (Fig. 3a, lines 2 and 6). Presence of a band of monomer size indicates that a fraction of Tcast1a is organized as a homogenous tandem array. A prominent band of dimer size in *ScaI* and *Eco47III* digestions suggests however that Tcast1a is interspersed with Tcast1b, and that such interspersed pattern represents the prevalent type of organization of TCAST satellite in *T. castaneum* genome.

In order to analyse specific organization of Tcast1b we performed Southern hybridization using a probe for the Tcast1b subfamily (Fig. 3b). Digestions with restriction enzyme *SatI* that cuts once within Tcast1b revealed presence of monomer band (Fig. 3b, line 1), which suggests existence of a fraction of Tcast1b in the form of homogenous arrays. However, prominent multimers relative to the monomer band in *SatI* digestion indicate the interspersed organization of Tcast1b as a prevalent type of organization. Such an interspersed pattern of organization of Tcast1b and Tcast1a is further confirmed by an almost identical pattern of *Eco47III* digestion obtained by mixed probes (Fig. 3a, line 2) and a Tcast1b specific probe (Fig. 3b, line 2). Double digestion with restriction enzymes *Eco47III* + *RsaI* that both have single recognition site in Tcast1a revealed a strong band of dimer size, pointing also to the interspersed pattern of Tcast1a and Tcast1b.

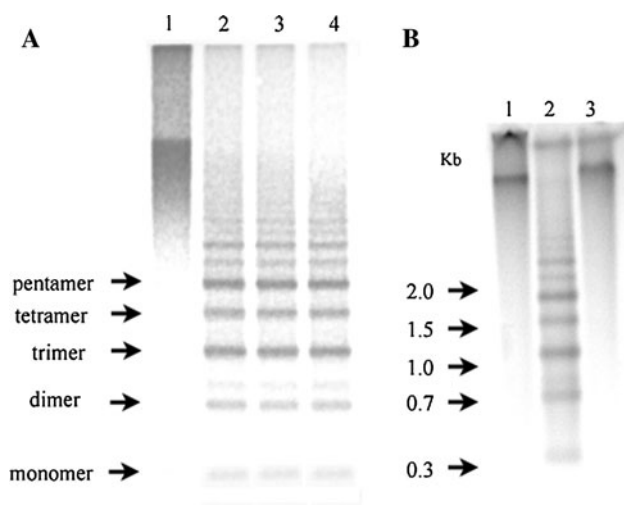
Southern hybridization analysis (Fig. 3a) revealed a difference in restriction profiles among enzymes *ScaI* and *Eco47III* that have a single recognition site in Tcast1a

**Fig. 3** Organization and variability of TCAST satellite within GA-2 strain of *T. castaneum*. Southern hybridization analysis was performed **a** using a mixture of biotin labelled Tcast1a and Tcast1b monomers as a hybridization probe. DNA was digested with: *AatII* (line 1), *Eco47III* (line 2), *NsiI* (line 3), *PaeI* (line 4), *PstI* (line 5) and *ScaI* (line 6). **b** Southern hybridization analysis using probe specific for Tcast1b subfamily. DNA was digested with *SatI* (line 1), *Eco47III* (line 2) and *Eco47III* + *RsaI* (line 3)



sequence (Fig. 2). In the *ScaI* hybridization profile, the dimer is the most prominent band as expected according to the highly conserved sequence at this restriction site and due to the interspersed organization of *Tcast1a* and *Tcast1b*. Differently, in *Eco47III* digestion, trimer and pentamer are the prevalent bands, indicating organization of the satellite into more complex higher order repeat units. Such a difference in profiles among the two enzymes could be explained by differences in sequence variability between the two restriction sites. Genomic sequence and most common sequence reveal the presence of a highly variable site (R) within the *Eco47III* recognition sequence (Fig. 2). According to the genomic consensus sequence, the variability occurs specifically within GCGC sequence present within the recognition site for *Eco47III* (AGCGCT). The predicted variability was confirmed by Southern hybridization after digestion of genomic DNA with restriction enzymes that recognize specifically this site: *Eco47III*, *HaeII* (recognition site RGCGCY) and *Hin6I* (recognition site GCGC) (Fig. 4a). The complete identity of digestion profiles obtained by the three enzymes demonstrates that the variability is restricted to GCGC sequence, as predicted by genomic sequencing.

An additional factor contributing to the specific variability in GCGC site could be methylation at a CpG site that blocks all three tested restriction enzymes. Digestion with *GlaI*, a restriction enzyme that cuts specifically methylated GCGC sequence (G5metCGC), demonstrated the absence of methylation at this site (Fig. 4b) which is in accordance with the previously reported absence of methylation in the



**Fig. 4** Southern hybridization analysis of *T. castaneum* genomic DNA from strain GA-2, using as a hybridization probe mixture of biotin labelled *Tcast1a* and *Tcast1b* monomers. DNA was digested with: **a** *Eco47III* (line 2, recognition site AGCGCT), *Hin6I* (line 3, recognition site GCGC), *HaeII* (line 4, recognition site PuGCGCPy); **b** *Hin6I* (line 2) and *GlaI* (line 3, recognition site G5metCGC). Lines 1 represent undigested *T. castaneum* DNA

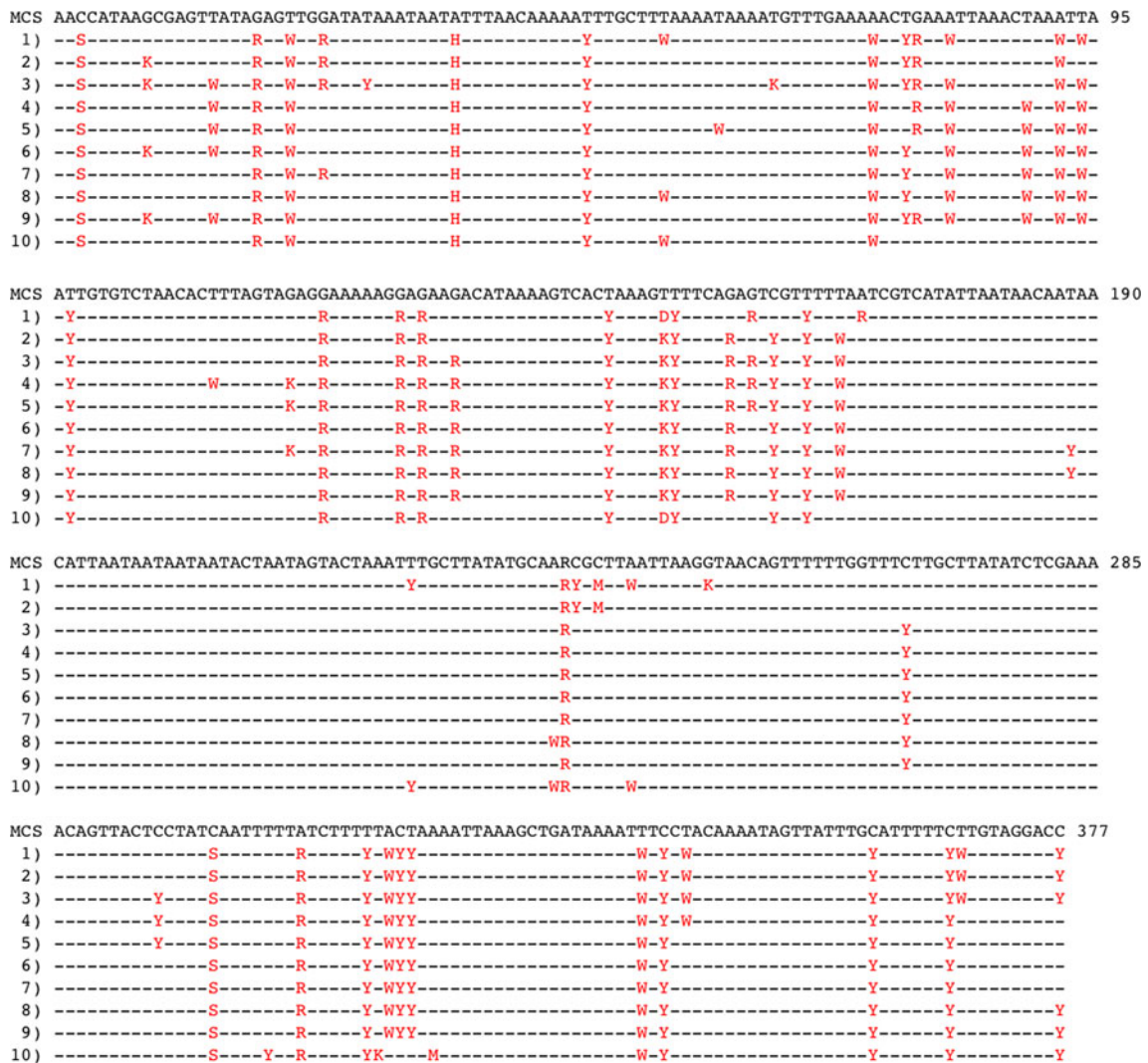
*T. castaneum* genome (Zemach et al. 2010). Southern hybridization revealed that variable GCGC sites are not randomly distributed within the tandem satellite array, but regularly phased within a subset of the satellite. Prominent satellite trimers and pentamers observed in digestions with *Eco47III*, *HaeII* and *Hin6I* indicate presence of a more complex organization of TCAST satellite units in the form of higher order repeats.

#### TCAST satellite variability among *T. castaneum* strains

To characterize the overall sequence variability of TCAST satellite DNA in the species *T. castaneum*, the genomic consensus sequences for TCAST satellite were obtained from nine other laboratory strains/populations, using the same pairs of primers. Distribution of the major variable sites and types of variability at these sites was very similar among all strains, as demonstrated for *Tcast1a* subfamily (Fig. 5). Usually only two bases exchange at the same position and such a pattern was preserved in most of populations. Although the most common sequences (MCS) of both *Tcast1a* and *Tcast1b* subfamilies were identical in all populations and correspond to sequences presented at Fig. 2, the genomic consensus sequences differed in the percentage of the minor bases present at some variable positions (Fig. 5). For example, in the laboratory strain from the CSL, at position 236 of *Tcast1a*, in addition to a major peak corresponding to the nt A, a T nt peak appears which corresponds in intensity to approximately 20% of the main peak. Since this mutation creates two new restriction sites for *NsiI* (ATGCAT) and *PaeI* (GCATGC) (Fig. 2), its presence can be confirmed using Southern hybridization analysis (Fig. 6). In all other strains except strain 61 where T makes approximately 15% of the main peak, T is present below 10%, while in GA-2 strain it is almost absent at position 236, as is confirmed by Southern hybridization (Fig. 3a).

Southern hybridization analysis of DNA from the CSL strain digested with the restriction enzymes *AatII*, *Eco47III*, *NsiI*, *PaeI*, *PstI* and *ScaI* and hybridized under the same conditions as GA-2 DNA (Fig. 3a), revealed very similar satellite profiles relative to the GA-2 strain. The only difference was in *NsiI* and *PaeI* digestions where a clear ladder pattern is visible in the CSL strain (Fig. 6, lines 3–4). This confirms the partial replacement of A at position 236 of *Tcast1a* in GA-2 strain with T in the CSL strain, as predicted by direct genomic sequencing.

We also checked by Southern hybridization if the pattern of variability observed at GCGC sequence that corresponds to the *Hin6I* restriction site is conserved among all tested strains (Fig. 7). Based on genomic consensus sequences (Fig. 5) a major variable site (R) at position 237 in *Tcast1a* is preserved in all strains and overlaps the *Hin6I*



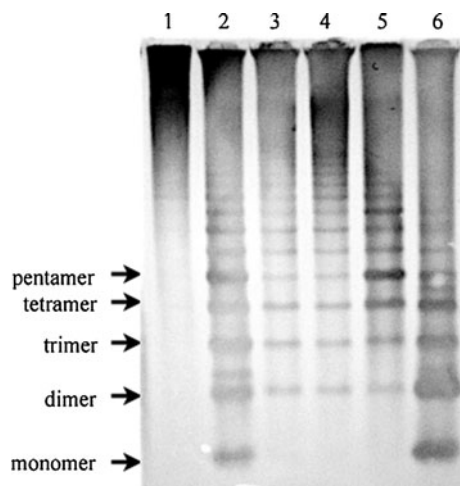
**Fig. 5** Alignment of the genomic consensus sequences (GCS) of Tcast1a from ten *T. castaneum* strains/populations: (1) GA-2, (2) 43, (3) 50, (4) 51, (5) 52, (6) 55, (7) 57, (8) 61, (9) ZAG-wild (10) CSL. Genomic consensus sequences are aligned relative to the most

common sequence (MCS) of Tcast1a that is conserved in all populations. Changes in the genomic consensus sequences relative to the most common sequence are indicated, while *dashed lines* indicate identical nucleotides

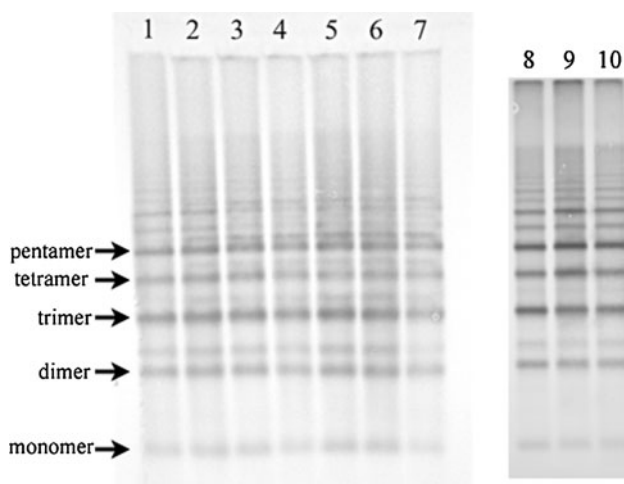
recognition site. Southern hybridization analysis using mixed probes of Tcast1a and Tcast1b reveals very similar variability at this site among ten strains. The hybridization pattern also indicates the same type of organization of TCAST satellite in all tested strains, as well as very similar amount of this satellite DNA in all strains (Fig. 7). Hybridization of the same blot using probes specific for either the Tcast1a or Tcast1b subfamily revealed almost identical restriction profiles, confirming an interspersed organization of Tcast1a and Tcast1b subunits as the prevalent type of satellite organization in all strains (not shown). In addition to main hybridization ladder composed of Tcast1a and Tcast1b subfamilies, existence of another ladder composed of weak, intermediate size bands is also visible (Fig. 7). Such a ladder could indicate presence

of a third TCAST satellite subfamily within *T. castaneum* genome.

We used dot blot hybridization with the probes specific for Tcast1a and Tcast1b to determine the relative amount of the two subfamilies in *T. castaneum* genome. The amount of Tcast1b in strain GA-2 is estimated to 15–17% of total DNA matching the previously determined amount of 17% for TCAST major satellite (not shown). Using the hybridization probe specific for Tcast1a, the amount of Tcast1a is estimated to be almost the same as those of Tcast1b, corresponding to approx. 20% of the genomic DNA in the all strains. This quantification shows that the newly described Tcast1a represents a highly abundant subfamily of TCAST satellite and that the total amount of TCAST satellite in the genome is much higher than



**Fig. 6** Organization and variability of TCAST satellite DNA within CSL strain of *T. castaneum*. Southern hybridization analysis was performed using a mixture of biotin labelled Tcast1a and Tcast1b monomers as a hybridization probe. DNA was digested with: *Aat*II (line 1), *Eco*47III (line 2), *Nsi*I (line 3), *Pae*I (line 4), *Pst*I (line 5) and *Sca*I (line 6)



**Fig. 7** Organization of TCAST satellite in ten strains/populations of *T. castaneum*. Southern hybridization analysis was performed with a mixture of biotin labelled Tcast1a and Tcast1b monomers. The same amount (1  $\mu$ g) of DNAs from the following strains/populations was digested with *Hin*6I restriction enzyme: 50 (line 1), 51 (2), 52 (3), 55 (4), 57 (5), 61 (6), ZAG-wild (7) CSL (8), 43 (9), GA-2 (10)

previously determined, ranging approximately between 35 and 40% of the whole genome.

## Discussion

Tandemly repeated satellite DNAs are subject to gene conversion, unequal crossing-over, transposition, slippage replication and other exchanges (Dover 2002). Because

mutations that change the sequence of one repeat are less common than recombination-induced replacement of one repeat by another, the repeats resemble each other much more than they would if they had been evolving independently. This phenomenon is known as concerted evolution. Because concerted evolution is usually unbiased and random, every version of repeat has an equal probability of being the one that replaces the others (Malik 2009). This simple mutation-recombination balance model, where recombination is the underlying force that homogenizes satellite repeats, does not seem to be able to completely explain the observed features of satellite DNAs (Ugarković 2005, 2009). Several studies have also pointed out the inadequacy of mutation and recombination alone to explain variability of tandem repeats suggesting that selection must play a role in their evolution (Walsh 1987; Stephan 1989; Charlesworth et al. 1994; Stephan and Cho 1994; Mravinac et al. 2005; Feliciello et al. 2006).

Subfamilies of satellite DNAs or different satellites can coexist within a genome forming a so called satellite library (Meštrović et al. 1998; Ugarković 2009). In the case of *T. castaneum*, two major abundant satellite subfamilies Tcast1a and Tcast1b are discovered that together make up 35% of the genome. Southern hybridization analysis confirmed the structures of both subfamilies obtained by direct genomic sequencing and revealed their mutual organization. The approach of direct sequencing of genomic PCR products used in this study does not exclude existence of other, less abundant subfamilies of TCAST. Despite limited number of TCAST satellite clones present in GenBank, using Blast search with the most common sequences of Tcast1a and Tcast1b as query sequences, it is possible to retrieve numerous clones highly homologous to each subfamily. This suggests abundance of the two subfamilies within the *T. castaneum* GenBank sequence database and *T. castaneum* genome, respectively. The two families are mutually interspersed and a composite Tcast1a + Tcast1b dimer seems to be the prevalent repeat unit. The pattern of digestion with some restriction enzymes indicates that Tcast1a and Tcast1b can form even more complex repeat units such as trimers and pentamers. Satellite subfamilies and higher order repeats are not often found in insects, in particular in beetles, although there are a few examples revealing the presence of complex dimers and trimers in some beetle species (reviewed in Palomeque and Lorite 2008). The interspersed pattern of two related satellites is characteristic for heterochromatin in the related species *T. madens* (Durajlija Žinić et al. 2000). However, the two satellites of *T. madens* are not regularly arranged like TCAST subfamilies, but form homogenous arrays of approx. 15 kb in length, composed of a single satellite, and those arrays are mutually interspersed within heterochromatin.



The sequences of the two TCAST satellite subfamilies are relatively homogenous and are characterized by major variable sites that are restricted to some positions and are preserved among ten strains from different geographical locations. However, differences in the percentage of minor bases present at some of the positions among *T. castaneum* strains were detected. In the case of the Tcast1a subfamily it is possible to follow partial replacement of the bases at a particular position among *T. castaneum* strains by restriction enzyme digestion. Such replacement and spreading of a new mutation represents the first step in the homogenization process and indicates the effect of concerted evolution on TCAST satellite DNA at the population level (Strachan et al. 1985). However, the major, predominant bases at all positions of either Tcast1a or Tcast1b subfamily were still identical in all strains. Satellite DNA sequence evolution at the population level was studied in different insect species such as cave crickets *Dolichopoda schiavazzi* (Bachmann et al. 1994), leaf beetle *Xanthogaleruca luteola* (Lorite et al. 2002), and *Drosophila buzzatti* species cluster (Kuhn et al. 2003; Kuhn and Sene 2004; de Franco et al. 2006). In all of these cases no specific mutations or other sequence features of satellite DNAs that could discriminate a population or groups of populations were discerned. Most of the analysed satellites exhibited high sequence homogeneity on both intra- and interpopulation levels, as revealed from the comparison of randomly cloned satellite repeats. Since all of these studies were based on the analysis of a relatively low number of cloned repeats that did not represent the whole population of repeats present within a genome, it is possible that slight changes in the mutational profiles remained hidden. Direct sequencing of PCR amplicons of a satellite DNA seems to represent a more suitable approach to study overall satellite DNA variability, as demonstrated here for TCAST satellite. To avoid bias in the estimation of variability due to primer specificity, different pairs of primers were combined and relaxed annealing conditions in PCR used. Comparison of genomic consensus sequences discerned discrete changes in the mutational profiles between populations and demonstrated the effect of concerted evolution during the early phases of satellite sequence divergence. Complete homogenization and fixation of population-specific mutations in satellite DNA was only observed in fishes (Elder and Turner 1994).

Expansions and contractions of satellite arrays within a library could also occur due to unequal crossingover and might contribute to continuous change of heterochromatin content and amount. Addition and loss of heterochromatic satellite DNA repeats is characteristic for strains of *Drosophila* and correlates with variation in the genome size (Bosco et al. 2007). Significant intraspecific heterogeneity of genome size was also detected in *Tribolium* species

(Alvarez-Fuster et al. 1991). However, our present study did not reveal change in the amount or organization of the two TCAST subfamilies at the population level. This indicates that the genome size heterogeneity is not related to this satellite DNA, although the change in the amount of some other repetitive DNAs could not be excluded. In conclusion, the analysis of major TCAST satellite DNA of beetle *T. castaneum* reveals that divergence of satellite DNA sequence characterized by a change in mutation profile occurs at the level of the population and precedes the evolution of species, representing the first step in the formation of a population-specific satellite DNA profile.

**Acknowledgments** This work was supported by EU FP6 Marie Curie Transfer of Knowledge Grant MTKD-CT-2006-042248 and grant 00982604 from the Croatian Ministry of Science. Isidoro Feliciello and Gianni Chinali were Marie Curie Fellows at Ruder Bošković Institute. We are grateful to Dani Dipietro for critical reading of the manuscript.

## References

- Alvarez-Fuster A, Juan C, Petitpierre E (1991) Genome size in *Tribolium* flour-beetles: inter- and intraspecific variation. *Genet Res* 58:1–5
- Bachmann L, Venanzetti F, Sbordoni V (1994) Characterization of a species-specific satellite DNA of *Dolichopoda schiavazzi* (Orthoptera, Rhaphidophoridae) cave crickets. *J Mol Evol* 39: 274–281
- Bosco G, Campbell P, Leiva-Neto JT, Markow TA (2007) Analysis of *Drosophila* species genome size and satellite DNA content reveals significant differences among strains as well as between species. *Genetics* 177:1277–1290
- Caraballo DA, Belluscio PM, Rossi MA (2010) The library model for satellite DNA evolution: a case study with the rodents of the genus *Ctenomys* (Octodontidae) from the Ibera marsh, Argentina. *Genetica* 138:1201–1210
- Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371:215–220
- de Franco FF, Kuhn GC, Sene FM, Manfrin MH (2006) Conservation of pBuM-2 satellite DNA sequences among geographically isolated *Drosophila gouveai* populations from Brazil. *Genetica* 128:287–295
- Dover G (2002) Molecular drive. *Trends Genet* 18:587–589
- Durajlija Žinić S, 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
- Elder FJ, Turner BJ (1994) Concerted evolution at the population level: pupfish *HindIII* satellite DNA sequences. *Proc Natl Acad Sci USA* 91:994–998
- Ellingsen A, Slamovits Ch, Rossi MS (2007) Sequence evolution of the major satellite DNA of the genus *Ctenomys* (Octodontidae, Rodentia). *Gene* 392:283–290
- Feliciello I, Picariello O, Chinali G (2005) The first characterization of the overall variability of repetitive units in a species reveals unexpected features of satellite DNA. *Gene* 349:153–164
- Feliciello I, Picariello O, Chinali G (2006) Intra-specific variability and unusual organization of the repetitive units in a satellite DNA from *Rana dalmatina*: molecular evidence of a new mechanism of DNA repair acting on satellite DNA. *Gene* 383:81–92

- Hinton HE (1948) A synopsis of the genus *Tribolium* Macleay with some remarks on the evolution of its species groups. Bull Entomol Res 39:13–55
- Juan C, Petitpierre E (1991) Chromosome number and sex-determining system in Tenebrionidae (Coleoptera). Advances in coleopterology. AEC Barcelona, pp 167–176
- Kuhn GC, Sene FM (2004) Characterization and interpopulation variability of a complex *HpaI* satellite DNA of *Drosophila seriema* (repleta group). Genetica 121:241–249
- Kuhn GC, Franco FF, Silva WA Jr, Martinez-Rossi NM, Sene FM (2003) On the pBuM189 satellite DNA variability among South American populations of *Drosophila buzzatii*. Hereditas 139:161–166
- Lorite P, Carrillo JA, Garneria I, Petitpierre E, Palomeque T (2002) Satellite DNA in the elm leaf beetle *Xanthogaleruca luteola* (Coleoptera, Chrysomelidae): characterization, interpopulation analysis and chromosomal location. Cytogenet Genome Res 98:302–307
- Malik H (2009) The centromere-drive hypothesis: a simple basis for centromere complexity. Prog Mol Subcell Biol 48:33–52
- 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, Ugarković Đ (2004) Conserved patterns in the evolution of *Tribolium* satellite DNAs. Gene 332:169–177
- Mravinac B, Plohl M, Ugarković Đ (2005) Preservation and high sequence conservation of satellite DNAs suggest functional constraints. J Mol Evol 61:542–550
- Palomeque T, Lorite P (2008) Satellite DNA in insects: a review. Heredity 100:564–573
- Picariello O, Feliciello I, Bellinello R, Chinali G (2002) S1 satellite DNA as a taxonomic marker in brown frogs: molecular evidence that *Rana graeca graeca* and *Rana graeca italica* are different species. Genome 45:63–70
- 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
- Richards S, Gibbs RA, Weinstock GM, Brown SJ, Denell R, Beeman RW, Gibbs R, Beeman RW, Brown SJ, Bucher G, Tribolium Genome Sequencing Consortium et al (2008) The genome of the model beetle and pest *Tribolium castaneum*. Nature 452:949–955
- Stephan W (1989) Tandem-repetitive noncoding DNA: forms and forces. Mol Biol Evol 6:198–212
- 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
- Ugarković Đ (2005) Functional elements residing within satellite DNAs. EMBO Rep 6:1035–1039
- Ugarković Đ (2009) Centromere-competent DNA: structure and evolution. Prog Mol Subcell Biol 48:53–76
- Ugarković Đ, Plohl M (2002) Variation in satellite DNA profiles—causes and effects. EMBO J 21:5955–5959
- 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
- Walsh JB (1987) Persistence of tandem arrays: implications for satellite and simple-sequence DNAs. Genetics 115:553–567
- Wang S, Lorenzen MD, Beeman RW, Brown SJ (2008) Analysis of repetitive DNA distribution patterns in the *Tribolium castaneum* genome. Genome Biol 9:R61
- Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328:916–919