In the red flour beetle *Tribolium castaneum* the major TCAST satellite DNA accounts for 35% of the genome and encompasses the pericentromeric regions of all chromosomes. Because of the presence of transcriptional regulatory elements and transcriptional activity in these sequences, TCAST satellite DNAs have also been proposed to be modulators of gene expression within euchromatin. Here, we analyze the distribution of TCAST homologous repeats in *T. castaneum* euchromatin and study their association with genes as well as their potential gene regulatory role. We identified 68 arrays composed of TCAST-like elements distributed on all chromosomes. Based on sequence characteristics the arrays were composed of two types of TCAST-like elements. The first type consists of TCAST satellite-like elements in the form of partial monomers or tandemly arranged monomers, up to tetramers, whereas the second type consists of TCAST-like elements embedded with a complex unit that resembles a DNA transposon. TCAST-like elements were also found in the 5′ untranslated region (UTR) of the CR1-3_TCa retrotransposon, and therefore retrotransposition may have contributed to their dispersion throughout the genome. No significant difference in the homogenization of dispersed TCAST-like elements was found either at the level of local arrays or chromosomes nor among different chromosomes. Of 68 TCAST-like elements, 29 were located within introns, with the remaining elements flanked by genes within a 262 to 404,270 nt range. TCAST-like elements are statistically overrepresented near genes with immunoglobulin-like domains attesting to their nonrandom distribution and a possible gene regulatory role.
certain species they account for the majority of genomic DNA, as in beetles from the coleopteran family Tenebrionidae (Ugarković and Plohl 2002). In the red flour beetle Tribolium castaneum, pericentromeric heterochromatin comprises approximately 40% of the genome, and TCAST satellite DNA has previously been characterized as the major satellite that encompasses centromeric as well as pericentromeric regions of all 20 chromosomes (Ugarković et al. 1996). TCAST satellite is composed of two subfamilies, Tcast1a and Tcast1b, which together comprise 35% of the whole genome. Tcast1a and Tcast1b have an average homology of 79% and are a similar size at 362 bp and 377 bp, respectively, but they are characterized by a divergent, subfamily specific region of approximately 100 bp (Feliciello et al. 2011). The genome sequencing project of T. castaneum has recently been completed (Richards et al. 2008). Sequencing involved the euchromatic portion of the genome, with >20% of the genome, corresponding to heterochromatic regions, excluded due to technical difficulties.

In this article, we searched for the presence of TCAST satellite-homologous elements within the assembled T. castaneum genome by using a comprehensive computational analysis. By searching the sequenced T. castaneum genome, we found 68 TCAST satellite DNA arrays within the euchromatin of all chromosomes. They were mapped to 5’ or 3’ ends, as well as within introns, of more than 100 protein-coding genes. Based on sequence characteristics, dispersed TCAST-like elements were classified into two groups. The first group includes partial TCAST satellite monomers or short arrays of tandemly arranged monomers up to tetramers. The second group contains TCAST-like element embedded within complex repeat units that contain two hallmarks of DNA transposons, terminal inverted repeats and target-size duplications. The evolutionary relationship and possible modes of dispersion of the two types of dispersed TCAST-like sequences are discussed. In addition, we examined the sequence divergence, phylogenetic relationship, and chromosomal distribution of the elements. Annotation, characterization, and classification of genes within the region of TCAST-like elements are reported, with the preferential localization of TCAST-like elements near specific groups of genes identified. Our results demonstrate for the first time, the enrichment of satellite DNA-like elements in the vicinity of genes with immunoglobulin-like domains and suggest their possible gene-regulatory role.

MATERIALS AND METHODS
BLASTN version 2.2.22+ was used to screen the NCBI refseq_genomic database of T. castaneum. All scaffolds that have not been mapped to linkage groups were also screened. The program was optimized to search for highly similar sequences (megablast) to the query sequence [TCAST consensus sequence (Ugarković et al. 1996)]. Genes flanking TCAST-homologous elements were found automatically by NCBI blast. Sequences corresponding to hits, as well as their flanking regions, were analyzed by dot plot (http://www.vivo.colostate.edu/molkit/dnadot/), using standard parameters (window size 9, mismatch limit 0), or more relaxed conditions (window size 11, mismatch limit 1), to determine the exact start and end site of specific TCAST-like elements. The TCAST transposon-like elements were analyzed in detail for the presence of hallmarks such as terminal inverted repeats (TIRs) and target-site duplications with the aid of the Gene Jockey sequence analysis program (for Apple Macintosh). Secondary structures were determined using the default parameters of the MFOLD program available online [http://mfold.rna.albany.edu/?q=mfold] (Zuker 2003)]. AT content was analyzed using BioEdit Sequence Alignment Editor (Hall 1999). Repbase, a reference database of eukaryotic repetitive DNA, was screened using WU-BLAST (Kohany et al. 2006).

Sequence alignment was performed using MUSCLE algorithm (Edgar 2004) combined with manual adjustment. All sequences were included in the alignment, with the exception of the ones that did not at least partially overlap with other sequences. Gblocks was used to eliminate poorly aligned positions and divergent regions of the alignments (Talavera and Castresana 2007). Alignments (original fasta files) are available upon request. jModelTest 0.1.1 software (Posada 2008) was used to infer best-fit models of DNA evolution—TPM3uf+G for transposon-like and A type elements and TPM1uf for B type elements. Maximum likelihood (ML) trees were estimated with the PhyML 3.0 software (Guindon and Gascuel 2003) using best-fit models. Markov chain Monte Carlo Bayesian searches were performed in MrBayes v. 3.1.2. (Huelsenbeck and Ronquist 2001) under the best-fit models (two simultaneous runs, each with four chains; 3 × 10⁶ generations; sampling frequency one in every 100 generations; majority rule consensus trees constructed based on trees sampled after burn-in). Branch support was evaluated by bootstrap analysis (1000 replicates) in ML and by posterior probabilities in Bayesian analyses. Pairwise sequence diversity (uncorrected P) was calculated using the MEGA 5.05 software (Tamura et al. 2011).

T. castaneum gene homologs in Drosophila melanogaster were searched using the OrthoDB Phylogenomic database. Each gene has OrthoDB identifier, with Uniprot data linked to OrthoDB (Waterhouse et al. 2011). To find sets of biological annotations that frequently appear together and are significantly enriched in a set of genes located near TCAST-like elements, program GeneCodis 2.0 available online (http://genecodis.dacya.ucm.es/) was used. GeneCodis generates statistical rank scores for single annotations and their combinations. To find all the possible combinations of annotations, Genecodis uses the apriori algorithm introduced by Agrawal et al. 1993. Once the annotations were extracted, a statistical analysis based on the hypergeometric distribution or the χ² test of independence was executed to calculate the statistical significance (P values) for each individual annotation or co-annotations.

Two-tailed hypergeometric test with Bonferroni correction (alpha = 0.025) was used to analyze the distribution of TCAST-like elements among T. castaneum chromosomes. In each chromosome the frequency of TCAST-like elements was compared with the frequency in the complete sample and the significance of deviations was calculated.

RESULTS
Identification of dispersed TCAST-like elements
Using the consensus sequence of TCAST satellite DNA (Ugarković et al. 1996) as a query sequence, we screened the NCBI refseq_genomic database of T. castaneum with the alignment program BLASTN version 2.2.22+. The program was optimized to search for highly similar sequences (megablast) and blast hits on the query sequence were analyzed individually. Alignments were mapped regarding start and end site, chromosome number, and total length. When the distance between two alignments on the same chromosome was short, the genomic sequence was further analyzed by dot plot to identify any potential continuity between the two alignments. Only genomic sequences with at least 140 nt (40% of TCAST monomer length) of continuous sequence and >80% identity to the TCAST consensus sequence were considered for further analysis. The total number of dispersed TCAST-like elements was 68, with 36 elements flanked by genes at both 5’ and 3’ ends, 3 elements flanked by a single gene either at 5’ or 3’ end (sequences no. 36, 39, 50), and the 29 elements positioned within introns (Table 1). Except 68 TCAST-like elements associated with genes, no other dispersed TCAST-like
elements were found within the assembled \textit{T. castaneum} genome. Analysis of scaffolds that have not been mapped to linkage groups revealed the presence of an additional 41 TCAST-like elements, but because they were not mapped to \textit{T. castaneum} genome and could possibly derive from heterochromatin, we did not consider them for further analysis.

There were only three cases in which two different TCAST-like elements were associated with the same gene: gene D6X2C4 contains TCAST-like sequences no. 6 and 13 within introns, gene D6X2U7 is flanked at 5’ and 3’ end by sequences no. 5 and 7, respectively, whereas gene D6WB29 is located at 3’ end of the sequence no. 53 and has sequence no. 52 within an intron. All other TCAST-like elements were positioned near or within different genes. Thus in total, there were 101 genes found in the vicinity of TCAST-like elements. Characteristics of the genes associated with TCAST-like elements, including gene identity number, gene name and chromosomal location, position relative to the associated TCAST-like element, and distances between TCAST-like elements and genes, are shown in Table 1 Distances between TCAST-like elements and genes range from 262 nt (gene positioned at 3’ site of the sequence no. 36), to a maximal distance of 404,270 nt (gene positioned at 5’ site of the sequence no. 5).

### Characteristics of TCAST-like elements

**TCAST satellite-like elements:** Sequence analysis of the 68 TCAST-like elements identified within the vicinity of genes enabled their classification into two groups. The first group contains partial TCAST satellite monomers or tandemly arranged elements, either complete or partial dimers, trimers, or tetramers (Table 1). The minimal size of satellite repeat was 203 nt (0.6 of complete TCAST monomer; sequence no. 15), whereas the maximal size was 1440 nt (four complete TCAST monomers; sequence no. 43; Table 1). In many sequences, two subtypes of TCAST satellite monomers were mutually interspersed: Tcast1a and Tcast1b. Tcast1b corresponds to the TCAST satellite consensus that was used as a query sequence (Ugarković et al. 1996), and Tcast1a corresponds to the TCAST subfamily described in Feliciello et al. 2011. Tcast1a and Tcast1b have an average homology of 79% and are of similar sizes at 362 bp and 377 bp respectively, but are characterized by a divergent, subfamily specific region of approximately 100 bp (Feliciello et al. 2011). There were 34 TCAST satellite-like elements found within or in the vicinity of 53 genes. Lengths of TCAST satellite-like elements (Table 1), their exact start and end sites within genomic sequence and composition (supporting information, Table S1) are provided.

To see whether there is any clustering of sequences of TCAST satellite-like elements due to the difference in the homogenization at the level of local array, chromosome, or among different chromosomes, sequence alignment and phylogenetic analysis were performed. Tcast1a and Tcast1b subunits were extracted from TCAST satellite-like sequences and analyzed separately. Alignment was performed on 24 Tcast1a subunits, ranging in size from 136 and 377 bp (File S1). The average pairwise distances between Tcast1a subunits of TCAST satellite-like sequences was 5.8%. Alignment adjustment using Gblocks, which eliminates poorly aligned positions and divergent regions, resulted in few changes; therefore, the original, unadjusted alignment was used for the construction of phylogenetic trees. Because the sequences differ in lengths and comprise regions of divergent variability, methods that take into account specific models of DNA evolution were considered as the most suitable for the construction of phylogenetic trees, ML and Bayesian (Markov chain Monte Carlo). The ML tree showed weak resolution with no significant support for clustering of sequences derived from the same satellite-like array or from the same chromosome. Similarly, the Bayesian tree demonstrated no significant sequence clustering (Figure 1A).

Alignment of 28 Tcast1b subunits, ranging from 159 bp to 363 bp (File S2), was also not significantly affected by adjustment with Gblocks; therefore, the unadjusted alignment was used for the construction of phylogenetic trees (Figure 1B). The average pairwise divergence between Tcast1b subunits, of TCAST satellite-like sequences, was 6.7%. With the ML phylogenetic tree, four groups composed of two or three sequences, were resolved by relatively low bootstrap values. However, the majority of Tcast1b subunit sequences remained unresolved. There was no clustering of subunits derived from the same array or the same chromosome (Figure 1B). Bayesian tree analysis produced one significantly supported cluster composed of 10 sequences derived from 7 chromosomes (Figure 1B).

**TCAST transposon-like elements:** The second group of TCAST-like repeats is represented by a complex element that contains an almost complete TCAST (or Tcast1b) monomer, and a TCAST monomer segment of approximately 121 bp in an inverted orientation. These two TCAST segments are separated by a nonsatellite sequence of approximately 306 bp. Both TCAST segments are part of TIRs that are approximately 269 bp long (Figure 2). As a result of the long TIRs, these elements are likely to form stable hairpin secondary structures and therefore resemble transposons. The nonsatellite part of sequence, common for all TCAST transposon-like elements, is unique in that it does not exhibit significant homology to any other sequence within the \textit{T. castaneum} genome. There were 34 TCAST transposon-like elements found within or in the vicinity of 50 genes. Their lengths (Table 1) and exact start and end sites within genomic sequence (Table S1) are provided. Sequence analysis of TCAST transposon-like elements determined that 13 of them were > 1000 bp, with a maximal size of 1181 bp (Table 1). The remaining TCAST transposon-like elements were shorter, with a minimal size of 314 bp (sequence no. 27), and usually lacking part of, or one or both, TIRs. Conserved TIRs are necessary for transposition, and if they are absent, truncated, or mutated so that the transposase cannot interact with the transposon sequence, the transposon cannot be mobilized and therefore represents a molecular fossil of a once active transposon (Capy et al. 1998). Despite mutations and partial truncations of TIRs within the TCAST transposon-like elements, and likely because of the length of the TIRs, most of the elements still preserve a stable secondary structure and could potentially remain mobile.

Some TCAST transposon-like elements >1000 bp have a 3-bp duplication at the site of insertion in the form of ACT. One TCAST transposon-like element (sequence no. 39) is inserted into another repetitive DNA, indicated as Tcast2, which had previously identified bioinformatically (Wang et al. 2008). Sequence analysis of this transposon-like element confirms the continuity of Tcast2 from the duplication site “ACT.” Typically, the size of target-site duplication is a hallmark of different superfamilies of eukaryotic DNA transposons, with \textit{mariner}/\textit{Tc1}, the only superfamily whose members are characterized by either 2- or 3-bp target-site duplication (Capy et al. 1998; Kapitonov and Jurka 2003; Feschotte and Pritham 2007). There are three open reading frames (ORFs) within TCAST transposon-like sequences, but the resulting putative proteins are very short and do not share similarity with any other proteins (Figure 2). The elements therefore do not code for transposases and are considered nonautonomous. Using the whole TCAST transposon-like elements as a query
| Uniprot | Entrez | Gene Name | Chr | Sat_seq. | Position, bp | Distance, bp | DM Homolog | FBgn | Type | Length | Copies |
|--------|--------|------------|-----|----------|--------------|-------------|------------|------|------|--------|--------|
| D6WZP1 | 662564 | Altered disjunction | 9   | 1        | 5'           | 18,773      | Q9VEH1     | FBgn0000063 | Satellite | 734    | 2.0    |
| D6WZP3 | 662624 | Ras-related protein Rab-26 | 9   | 1        | 3'           | 7795        | Q9VP48     | FBgn0086913 | Satellite | 734    | 2.0    |
| D6WZL9 | 661947 | Probable serine/threonine-protein kinase | 9   | 2        | Inside       | 99,669      | Q8IP89     | FBgn000114 | Satellite | 976    | 2.0    |
| D6X226 | 660275 | Numb | 9   | 4        | 5'           | 1520        | Q9VE80     | FBgn0038610 | Satellite | 517    | 1.4    |
| D6X238 | 661741 | Dopamine receptor 1 | 9   | 6        | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X2D0 | 660440 | Short-chain dehydrogenase | 9   | 7        | 5'           | 51,101      | Q9VCYS     | FBgn0053110 | Satellite | 394    | 1.1    |
| D6X1E7 | 656884 | Cytochrome P450 306A1 | 9   | 8        | Inside       | 40,270      | Q9VWR5     | FBgn004959 | Satellite | 1058   | 2.9    |
| D6X2U7 | 656977 | Elongase | 9   | 9        | 3'           | 9947        | Q9VCY6     | FBgn0038986 | Satellite | 1058   | 2.9    |
| D6X2C4 | 660195 | Dopamine receptor 1 | 9   | 10       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X366 | 657055 | elongation of very long chain fatty acids protein | 9   | 11       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X0D7 | 657748 | Ret oncogene | 9   | 12       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X0E1 | 657829 | Dpr9 | 9   | 13       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X2H8 | 655561 | Elongase | 9   | 14       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X2U7 | 656977 | Elongase | 9   | 15       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X2C4 | 660195 | Dopamine receptor 1 | 9   | 16       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X0E1 | 657829 | Dpr9 | 9   | 17       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X2H8 | 655561 | Elongase | 9   | 18       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X2U7 | 656977 | Elongase | 9   | 19       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X2C4 | 660195 | Dopamine receptor 1 | 9   | 20       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X0E1 | 657829 | Dpr9 | 9   | 21       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X0E1 | 657829 | Dpr9 | 9   | 22       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X0E1 | 657829 | Dpr9 | 9   | 23       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X0E1 | 657829 | Dpr9 | 9   | 24       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X0E1 | 657829 | Dpr9 | 9   | 25       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X0E1 | 657829 | Dpr9 | 9   | 26       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X0E1 | 657829 | Dpr9 | 9   | 27       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X0E1 | 657829 | Dpr9 | 9   | 28       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X0E1 | 657829 | Dpr9 | 9   | 29       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X0E1 | 657829 | Dpr9 | 9   | 30       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X0E1 | 657829 | Dpr9 | 9   | 31       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X0E1 | 657829 | Dpr9 | 9   | 32       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X0E1 | 657829 | Dpr9 | 9   | 33       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X0E1 | 657829 | Dpr9 | 9   | 34       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| D6X0E1 | 657829 | Dpr9 | 9   | 35       | Inside       | 7128        | Q9VCY6     | FBgn0038986 | Satellite | 394    | 1.1    |
| Uniprot | Entrez | Gene Name | Chr | Sat_seq. | Position | Distance, bp | DM Homolog | FBgn | Type | Length | Copies |
|--------|--------|-----------|-----|----------|----------|-------------|------------|------|------|--------|--------|
| D6WEC2 | 664189 | Neprhin   | 3   | 32       | Inside   | Q9W4T9      | FBgn0028369 | Transposon | 666   |
| D6W96  | 100142620 | Heat shock protein 70 | 3 | 33 | In | P11147 | FBgn0001219 | Transposon | 1058 |
| D6WG02 | 664917 | N-acetylglycosaminyltransferase vi | 7 | 41 | Inside | Q9VUDH | FBgn0036446 | Transposon | 319   |
| D6WYD1 | 658691 | Putative uncharacterized protein | 8 | 35 | 5' | 385,712 | Q8SY79 | FBgn0032249 | Satellite | 625.17 |
| D6WYN3 | 658691 | serine-type protease inhibitor | 8 | 35 | 3' | 58,583 | Q9VSC9 | FBgn0035833 | Satellite | 625.17 |
| D6WYA2 | 658691 | Putative uncharacterized protein | 8 | 38 | Inside | Q7K3E2 | FBgn0003131 | Transposon | 582   |
| D6WYD1 | 658691 | Putative uncharacterized protein | 8 | 39 | 5' | 7165 | Q94534 | FBgn0013433 | Transposon | 1181 |
| D6WUX6 | 662235 | Putting on character protein | 8 | 40 | Inside | Q7KUK9 | FBgn0036454 | Transposon | 440   |
| D6XE01 | 654938 | defective ribosome extension response | 7 | 41 | Inside | Q9VFD9 | FBgn0038282 | Satellite | 722.20 |
| D6WYH8 | 662021 | Ribosome-releasing factor 2, mitochondrial | 7 | 4 | Inside | Q9VCX4 | FBgn003125 | Transposon | 905.25 |
| D2A2C2 | 663849 | Putative uncharacterized protein | 4 | 45 | 5' | 9489 | Q9V5S3 | FBgn0013300 | Satellite | 649.15 |
| D2A2C2 | 663849 | Putative uncharacterized protein | 4 | 46 | 3' | 9,450 | Q9W191 | FBgn0034994 | Satellite | 649.15 |
| D2A2C2 | 663849 | Putative uncharacterized protein | 4 | 46 | 5' | 58,200 | Q9S528 | FBgn0033786 | Satellite | 558.16 |
| D2A2C2 | 663849 | Putative uncharacterized protein | 4 | 46 | 3' | 7,000 | P8B591 | FBgn0012051 | Satellite | 558.16 |
| D2A2D2 | 660808 | Kinesin-like protein | 4 | 47 | Inside | Q9VLW2 | FBgn0003295 | Transposon | 508   |
| D2A2C2 | 660808 | Kinesin-like protein | 4 | 47 | Inside | Q9VHL1 | FBgn0037633 | Satellite | 337   |
| D6WB5 | 650258 | E74 | 2 | 49 | 5' | 60,525 | P20105 | FBgn0000567 | Satellite | 770.21 |
| D6WB73 | 654932 | organic cation transporter | 2 | 49 | 3' | 2638 | Q7K3M6 | FBgn0034467 | Satellite | 770.21 |
| D6WGB8 | 659129 | pre-mRNA-splicing helicase BR2 | 2 | 50 | 3' | 4811 | Q9VUV9 | FBgn0036548 | Satellite | 770.21 |
| D6WB14 | 658844 | monophenolic amine tyramine | 2 | 51 | 3' | 7,995 | P22250 | FBgn0004514 | Transposon | 567.09 |
| D6WB15 | 655798 | Cuticular protein 47Ef | 2 | 51 | 3' | 16,739 | A1Z8H7 | FBgn0033608 | Transposon | 567.09 |
| D6WB29 | 657778 | Endoprotease FURIN | 2 | 52 | Inside | P30432 | FBgn0004598 | Transposon | 1045.23 |
| A8D1V5 | 657942 | Nicotinic acetylcholine receptor subunit alpha11 | 2 | 52 | 3' | 13,645 | P25162 | FBgn0004118 | Transposon | 1021.31 |

(continued)
To test whether there is any chromosome-specific sequence clustering of TCAST transposon-like sequences that could suggest difference in homogenization within chromosome and among different chromosomes, the alignment and subsequent phylogenetic analysis of TCAST transposon-like sequences was performed. Because TCAST transposon-like elements differ significantly in size (314–1181 nt), the alignment and phylogenetic analyses were performed on 25 elements that mutually overlap in their sequences, whereas the other nine TCAST transposon-like elements were excluded from the analysis due to the very low overlapping with other elements. Alignment was additionally adjusted using Gblocks (File S3). The average pairwise divergence among TCAST transposon-like sequences was 12.7%. ML and Bayesian methods gave similar tree topologies (Figure 1C). The ML tree showed very weak resolution of TCAST transposon-like sequences was 12.7%. The average pairwise divergence among TCAST transposon-like sequences was 12.7%. The ML and Bayesian methods gave similar tree topologies (Figure 1C). The ML tree showed very weak resolution of TCAST transposon-like sequences was 12.7%. The average pairwise divergence among TCAST transposon-like sequences was 12.7%. The ML and Bayesian methods gave similar tree topologies (Figure 1C). The ML tree showed very weak resolution of TCAST transposon-like sequences was 12.7%. The average pairwise divergence among TCAST transposon-like sequences was 12.7%. The ML and Bayesian methods gave similar tree topologies (Figure 1C). The ML tree showed very weak resolution of TCAST transposon-like sequences was 12.7%. The average pairwise divergence among TCAST transposon-like sequences was 12.7%. The ML and Bayesian methods gave similar tree topologies (Figure 1C). The ML tree showed very weak resolution of TCAST transposon-like sequences was 12.7%. The average pairwise divergence among TCAST transposon-like sequences was 12.7%. The ML and Bayesian methods gave similar tree topologies (Figure 1C). The ML tree showed very weak resolution of TCAST transposon-like sequences was 12.7%. The average pairwise divergence among TCAST transposon-like sequences was 12.7%. The ML and Bayesian methods gave similar tree topologies (Figure 1C). The ML tree showed very weak resolution of TCAST transposon-like sequences was 12.7%. The average pairwise divergence among TCAST transposon-like sequences was 12.7%. The ML and Bayesian methods gave similar tree topologies (Figure 1C). The ML tree showed very weak resolution of TCAST transposon-like sequences was 12.7%. The average pairwise divergence among TCAST transposon-like sequences was 12.7%. The ML and Bayesian methods gave similar tree topologies (Figure 1C). The ML tree showed very weak resolution of TCAST transposon-like sequences was 12.7%. The average pairwise divergence among TCAST transposon-like sequences was 12.7%. The ML and Bayesian methods gave similar tree topologies (Figure 1C). The ML tree showed very weak resolution of TCAST transposon-like sequences was 12.7%. The average pairwise divergence among TCAST transposon-like sequences was 12.7%. The ML and Bayesian methods gave similar tree topologies (Figure 1C).

**Distribution of TCAST-like elements on T. castaneum chromosomes**

TCAST-like elements found in the vicinity of genes were distributed on all 10 T. castaneum chromosomes (Table 1). Positions of constitutive heterochromatin and euchromatin were assigned on the haploid set of T. castaneum chromosomes, based on C-banding data (Stuart and Mocellin 1995) and Tribolium castaneum 3.0 Assembly data (Figure 3). Within euchromatic segments, the position of each TCAST-like element is specifically indicated (Figure 3) based on the position within the genomic sequence (Table S1). TCAST-like elements were dispersed on both arms of chromosomes 3, 5, 9, and 7, whereas on other chromosomes they were located on a single arm (Figure 3). The number of TCAST-like elements ranged from 2 on chromosome 1(X) to 17 on chromosomes 3 and 9. To detect whether TCAST-like elements were distributed randomly among the T. castaneum chromosomes or whether there was a significant over or underrepresentation of the elements on some chromosomes we performed hypergeometric distribution analysis test. The analysis revealed no statistically significant deviation in the number of TCAST-like elements among the chromosomes (Figure S1), pointing to their random distribution.

To determine whether there was a target preference for the insertion of TCAST-like elements, for example high AT content or another sequence characteristic, we analyzed the AT content within 100 bp of the flanking regions for each TCAST-like element, from both 5’ and 3’ sites (Figure S2 and Figure S3). The average AT content
of the flanking regions for both TCAST satellite-like elements and TCAST transposon-like elements did not differ significantly from the average AT content of the whole T. castaneum genome or from the AT content of randomly selected intergenic regions and introns. Thus, this finding suggests that with regard to AT content, there is no target preference for the insertion of TCAST-like elements. Furthermore, alignment and comparison of all flanking sequences of TCAST-like elements did not identify any common sequence motifs.

**Genes in the vicinity of TCAST-like elements**

Uniprot gene numbers were used as identifiers of genes located in the vicinity of TCAST-like elements (gene names shown in Table 1). Uniprot gene numbers for homologous genes found in Drosophila melanogaster are also indicated (Table 1). Detailed description of the genes, including molecular function of their protein products, biological processes in which these proteins are involved, and their cellular localization (cellular component), are shown (Table S1). Each identified gene is assigned to a particular TCAST-like element within its vicinity, and the precise position of TCAST-like elements in genomic sequence (start and end site) is indicated (Table S1). Functional analysis revealed that 17 of 101 genes correspond to putative uncharacterized proteins, whereas the remaining genes are involved in different molecular functions and diverse biological processes. Among the proteins, a proportion is characterized by ATP binding activity (13 proteins) and involvement in protein phosphorylation and/or signal transduction (9 proteins; Table S1).

To determine whether TCAST-like elements are distributed randomly relative to genes or whether they are overrepresented near specific groups of genes, we used GeneCodis 2.0 to provide a statistical representation of the genes associated with TCAST-like elements. Because many genes are still not annotated in T. castaneum and furthermore T. castaneum genomic data are not included in GeneCodis, we used gene numbers for orthologous genes from D. melanogaster for the analysis and compared them with the whole set of 14,869 genes annotated in D. melanogaster. GeneCodis analysis revealed that TCAST-like elements are located near nine genes characterized as members of the immunoglobulin protein superfamily. Because there are only 134 immunoglobulin-like genes present within the total set of D. melanogaster genes, random distribution of TCAST-like elements would result in their occurrence near approximately a single immunoglobulin-like gene. The presence of TCAST-like elements in the vicinity of nine immunoglobulin-like genes therefore represents a statistically significant overrepresentation (0.00000427). All nine genes exhibit structural features of immunoglobulin-like, immunoglobulin subtype 1 and immunoglobulin subtype 2 proteins and are associated with the following TCAST transposon-like elements: 25 at the 3’ end, 28 and 39 at the 5’ end, 32 and 40 within introns, and TCAST satellite-like elements: 8 at the 3’ end, 19 and 62 at the 5’ end, and 41 within intron (Table 1). A minimal distance between TCAST-like element and immunoglobulin-like gene was 7165 bp and a maximal 173,881 bp (Table 1). Molecular function of most of immunoglobulin-like genes is unknown, and they are involved in different biological processes such as cell adhesion, protein phosphorylation, and axon guidance (Table S1). Although all nine genes belong to immunoglobulin superfamily, they did not exhibit sequence similarity, which could suggest role of duplication in their evolution and spreading. The position of TCAST-like elements relative to the genes also was not consistent with the possibility that TCAST-like elements duplicated along with the immunoglobulin-like genes.

Overrepresentation of TCAST-like elements was also found near genes that exhibit ATP-binding activity and axon guidance properties but with a marginal significance (0.0183374 and 0.00865139). For the rest of genes, no significant overrepresentation of TCAST-like elements was detected. Thus, enrichment of TCAST-like elements in the vicinity of immunoglobulin-like genes potentially implicates a role of TCAST-like elements in the regulation of these genes.
**DISCUSSION**

TEs are classified in several dozen families based on transposition mechanisms and different dynamics properties (Hua-Van et al. 2005). Active TEs encode the enzymes necessary for their transposition, either to move between nonhomologous regions in the genome or to copy themselves to other positions. In many cases, TEs do not produce their own enzymes but are able to use those from functional copies or even from other TEs families. Defective and inactive TEs often are amplified in regions of low recombination such as heterochromatin and may form tandemly repeated satellite DNAs. The origin of satellite DNA array from transposon-like elements is reported for many insects such as *Drosophila melanogaster* (Agudo et al. 1999), *Drosophila guanche* (Miller et al. 2000), and the beetle *Misolampus goudoti* (Pons 2004) whereas the retroviral-like features were first observed in the satellite DNA from rodents of the genus *Ctenomys* (Rossi et al. 1993).

Transposons can be inserted into other repetitive sequences such as satellite DNAs, as has been observed for the *mariner*-like element and MITE element, both inserted into satellite DNA of the ant *Messor bouvieri* (Palomeque et al. 2006). Searching for repetitive elements homologous to the TCAST repeat within Repbase (http://www.girinst.org/repbase/) revealed that 5’ UTR of nonlong terminal repeat retrotransposon CR1-3_TCa (Jurka 2009c) shares a high similarity of 83% with a 444-bp long TCAST sequence composed of 1.2 tandem monomers (Figure 1). Other CR1 subfamilies identified within *T. castaneum* such as CR1-1_TCa, CR1-2_TCa, and CR1-4_TCa, published in Repbase, do not share similarity to CR1-3 and do not contain TCAST similar sequence. We propose that CR1-3 was inserted within TCAST satellite array and through recombination has acquired a part of TCAST sequence. Newly acquired TCAST element could act as a promoter because TCAST satellite DNA has an internal promoter for RNA Pol II (Pezer and Ugarkovic 2012) and becomes a new functional 5’ UTR. Subsequent retrotransposition of CR1-3_TCa could explain the dispersion of TCAST within the euchromatin (Figure 4). Three CR1-3_TCa elements with TCAST in the 5’ UTR were identified within scaffolds that have not been mapped to linkage groups. However, truncated fragments with partial homology to CR1-3_TCa retrotransposon can be mapped within *T. castaneum* genome, some of them in the vicinity of TCAST elements. Such arrangement also indicates the role of CR1-3_TCa in the spreading of TCAST elements. There is also a possibility that TCAST satellite DNA originates from CR1-3 retrotransposon which was, after inactivation, amplified within the heterochromatin region. In the case of TCAST transposon-like elements, part of the satellite sequence is incorporated within TIRs which are characteristic for DNA transposons. The presence of target-site duplications at the sites of insertions of some TCAST transposon-like elements also indicates transposition as a mode of spreading of TCAST.
elements. Parts of satellite DNA elements can be found within some transposons, such as pDe transposon (Evgen'ev et al. 1982; Zelentsova et al. 1986) whose long direct terminal repeats show significant sequence similarity to the pvB370 satellite DNA, located in the centromeric heterochromatin of a number of species of the Drosophila viridis group (Heilkinen et al. 1995). The presence of short stretches of PisTR-A satellite DNA sequences within 3' UTR of Ogre retrotransposons dispersed in the pea (Pisum sativum) genome was reported (Macas et al. 2009). Furthermore, the mobilization of subtelomeric repeats upon excision of the transposable P element from tandemly repeated subtelomeric sequences has been observed (Thompson-Stewart et al. 1994).

Incorporation of part of a TCAST satellite DNA sequence into a (re)transposable element, and its subsequent mobilization and spreading by (re)transposition, may explain the distribution of TCAST element in the vicinity of genes within euchromatin. Satellite DNA sequences are prone to undergo recurrent repeat copy number expansion and contraction in divergent lineages as well as among populations of the same species (Bosco et al. 2007). This amplification appears to be random and does not correlate with phylogeny of the species (Pons et al. 2004; Lee et al. 2005; Bulazel et al. 2007). Amplification of a satellite sequence is reported to occur as a result of unequal crossing over or duplicative transposition (Smith 1976; Ma and Jackson 2006). The discovery of human extrachromosomal elements originating from satellite DNA arrays in cultured human cells and different plant species indicates the possible existence of additional amplification mechanisms based on rolling-circle replication (Assum et al. 1993; Navrátilová et al. 2008). It has been proposed that satellite sequences excised from their chromosomal loci via intras Strand recombination could be amplified in this way, followed by reintegration of tandem arrays into the genome (Feliciello et al. 2006). Moreover, it is possible that such a mechanism affected TCAST satellite DNA, and that extrachromosomal circles of TCAST were reintegrated into different genome locations by homologous recombination based on short stretches of sequence similarity between TCAST satellite and target genomic sequence (Figure 4).

Integrated TCAST sequences are mainly composed of interspersed elements belonging to two major subfamilies, Tcast1a and Tcast1b, which is a prevalent type of organization in pericentromeric heterochromatin (Feliciello et al. 2011). This finding indicates that the origin of dispersed euchromatic TCAST elements may be duplication of heterochromatin copies.

The distribution of TCAST-like elements relative to protein coding genes revealed no specific preference for insertions within introns or at 5' or 3' ends of genes. TCAST-like elements are distributed on all chromosomes with no significant deviation in the number among the chromosomes, and phylogenetic analysis did not detect any significant sequence clustering of TCAST-like elements derived from the same chromosome. Dispersed TCAST satellite-like elements produce tandem arrays up to tetramers, but repeats from the same array do not reveal any significant clustering on phylogenetic trees. This finding indicates there is no significant difference in the homogenization of TCAST satellite-like repeats at the level of local arrays or chromosome or among different chromosomes. The average pair-wise sequence divergence (6% for dispersed TCAST satellite-like repeats) is greater than the usual divergence of satellite elements located in heterochromatin of tenebrionid beetles [approximately 2% (Ugarković et al. 1996)]. This difference in homogeneity between repeats located in heterochromatin and euchromatin may be explained by a lower rate of gene conversion affecting dispersed satellite-like elements or by a specific mechanism of DNA repair acting on satellite DNA (Feliciello et al. 2006). TCAST transposon-like elements dispersed among the genes within euchromatin have an even greater average sequence divergence (approximately 12%) and also exhibit no significant chromosome-specific sequence clustering, indicating a similar rate of homogenization within and among the chromosomes. Relatively high
sequence divergence of TCAST transposon-like elements and the significant truncation of the majority of them, indicates that the transposition of these elements did not occur very recently and that these elements could be considered as molecular fossils of the functional TCAST transposon-like elements.

Cis-regulatory elements, such as promoters or transcription factor binding sites, are predicted in some satellite DNAs (Pezer et al. 2011). Transcription from promoters for RNA Pol II is also characteristic for pericentromeric satellite DNAs from the beetles *Palorus rützelsorgei* and *Palorus subdepressus* (Pezer and Ugarkovic 2008, 2009). Temperature-sensitive transcription of TCAST satellite DNA from an internal RNA Pol II promoter has been demonstrated (Pezer and Ugarkovic 2012). Based on these findings, it can be proposed that TCAST elements located in the vicinity of genes may function as alternative promoters, and transcripts derived from them may interfere with the expression of neighboring genes. This type of regulation is often observed for retrotransposons positioned immediately 5′ of protein genes (Faulkner et al. 2009). In addition, some tissue-specific gene promoters are derived from retrotransposons (Ting et al. 1992; Samuelson et al. 1996). Because of rapid evolutionary turnover, satellite DNA sequences often are restricted to a group of closely related species, or in some instances are species specific. This is the case with TCAST satellite DNA, which is not even detected in the congeneric *Tribolium* species. If restricted satellite DNAs have regulatory potential, then insertion of these elements in vicinity of genes could contribute to the establishment of lineage-specific or species-specific patterns of gene expression. Annotation of genes in proximity to TCAST-like elements demonstrated a statistical overrepresentation of certain groups of genes, for example, those with immunoglobulin-like domains. Recently, in the fish *Salvelinus fontinalis*, a regulatory role of a 32-bp satellite repeat, located in an intron of the major histocompatibility complex gene (MHIIB), on MHIIB gene expression was demonstrated (Croisetiere et al. 2010). The level of gene expression depends on temperature, as well as the number of satellite repeats, and indicates a role for temperature-sensitive satellite DNA in gene regulation of the adaptive immune response. Further studies are necessary to determine whether TCAST-like elements exhibit a potential regulatory role on nearby genes. The transcriptional potential of satellite DNAs as well as their distribution close to protein-coding genes, as shown in this study, provides strong support, that in addition to transposons, satellite DNAs represent a rich source for the assembly of gene regulatory systems.

ACKNOWLEDGMENTS

This work was supported by Croatian Ministry of Science, Education and Sport (grant no. 098-0982913-2832), European Union FP6 Marie Curie Transfer of Knowledge (grant MTKD-CT-2006-042248), and COST Action TD0905 “Epigenetics: Bench to Bedside.”

LITERATURE CITED

Agrawal, R., T. Inseliński, and A. Swami, 1993  Mining association rules between sets of items in large databases, pp. 207–216 in the Proceedings of the ACM SIGMOD International Conference on Management of Data, edited by P. Buneman and S. Jajodia. ACM Press, New York.

Agudo, M., A. Losada, J. P. Abad, S. Pimpinelli, P. Ripoll et al., 1999  Centromeres from telomeress? The centromeric region of the Y chromosome of *Drosophila melanogaster* contains a tandem array of telomeric HeT-A- and TART-related sequences. Nucleic Acids Res. 27: 3318–3324.

Assum, G., T. Fink, T. Steinbeisser, and K. J. Fisel, 1993  Analysis of human extrachromosomal DNA elements originating from different beta-satellite subfamilies. Hum. Genet. 91: 489–495.

Bosco, G., P. Campbell, I. T. Leiva-Neto, and T. A. Markow, 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.

Britten, R. J., and E. H. Davidson, 1971  Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. Q. Rev. Biol. 46: 111–138.

Bulazel, K. V., G. C. Ferreri, M. D. Eldridge, and R. J. O’Neill, 2007  Species-specific shifts in centromere sequence composition are coincident with breakpoint reuse in karyotypically divergent lineages. Genome Biol. 8: R170.

Capy, P., C. Bazin, D. Higuet, and T. Langin, 1998  Dynamics and Evolution of Transposable Elements. Springer-Verlag, Austin, TX.

Croisetiere, S., L. Bernatchez, and P. Belhumeur, 2010  Temperature and length-dependent modulation of the MH class IIβ gene expression in brook charr (*Salvelinus fontinalis*) by a cis-acting minisatellite. Mol. Immunol. 47: 1817–1829.

Edgar, R. C., 2004  MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32: 1792–1797.

Ewen’ev, M. B., G. N. Yenikolopov, N. I. Peunova, and Y. V. Ilyin, 1982  Transposition of mobile genetic elements in interspecific hybrids of *Drosophila*. Chromosoma 85: 375–386.

Faulkner, G. J., Y. Kimura, C. O. Daub, S. Wani, C. Plessy et al., 2009  The regulated retrotransposon transcriptome of mammalian cells. Nat. Genet. 41: 563–571.

Feliciello, I., O. Picariello, and G. Chinali, 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.

Feliciello, I., G. Chinali, and D. Ugarkovic, 2011  Structure and evolutionary dynamics of the major satellite in the red flour beetle *Tribolium castaneum*. Genetica 139: 999–1008.

Feschotte, C., 2008  Transposable elements and the evolution of regulatory networks. Nat. Rev. Genet. 9: 397–405.

Feschotte, C., and E. J. Pritcham, 2007  DNA transposons and the evolution of eukaryotic genomes. Annu. Rev. Genet. 41: 331–368.

Guindon, S., and O. Gascuel, 2003  A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52: 696–704.

Hall, T. A., 1999  BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids Symp. Ser. 41: 95–98.

Heikinen, E., V. Launonen, E. Müller, and L. Bachmann, 1995  The pVB370 BamHI satellite DNA family of the *Drosophila viridis* group and its evolutionary relation to mobile dispersed genetic pDv elements. J. Mol. Evol. 41: 604–614.

Hua-Van, A., A. Le Rouzic, C. Maisonnhaute, and P. Capy, 2005  Abundance, distribution and dynamics of retrotransposable elements and transposons: similarities and differences. Cytogenet. Genome Res. 110: 426–440.

Hulsenbeck, J. P., and F. Ronquist, 2001  MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17: 754–755.

Jurka, J., 2009a  Mariner-1_TCa. Repbase Rep. 9: 674.

Jurka, J., 2009b  Mariner-2_TCa. Repbase Rep. 9: 675.

Jurka, J., 2009c  CR1–3_TCa. Repbase Rep. 9: 737.

Kapitonov, V. V., and J. Jurka, 2003  Molecular paleontology of transposable elements. Springer-Verlag, Austin, TX.

Kuhn, G. C., H. Küttler, O. Moreira-Filho, and J. S. Heslop-Harrison, 2012  The 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.

Kohany, O., A. J. Gentles, L. Hankus, and J. Jurka, 2006  Annotation, subfamilies and variability and dynamics of mobile DNA elements. J. Mol. Evol. 66: 999–1008.
Lowe, C. B., G. Bejerano, and D. Haussler, 2007  Thousands of human mobile element fragments undergo strong purifying selection near development genes. Proc. Natl. Acad. Sci. USA 104: 8005–8010.

Ma, J., and S. A. Jackson, 2006  Retrotransposon accumulation and satellite amplification mediated by segmental duplication facilitate centromere expansion in rice. Genome Res. 16: 251–259.

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

Miller, W. J., A. Nagel, J. Bachmann, and L. Bachmann, 2000  Evolutionary dynamics of the SGM transposon family in the Drosophila obscura species group. Mol. Biol. Evol. 17: 1597–1609.

Navrátilová, A., A. Kobližková, and J. Macas, 2008  Survey of extrachromosomal circular DNA derived from plant satellite repeats. BMC Plant Biol. 8: 90.

Palomeque, T., J. A. Carrillo, M. Munos-Lopez, and P. Lorite, 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.

Pezer, Z., and D. Ugarović, 2008  RNA Pol II promotes transcription of centromeric satellite DNA in beetles. PLoS ONE 3: e1594.

Pezer, Z., and D. Ugarović, 2009  Transcription of pericentromeric heterochromatin in beetles – satellite DNAs as active regulatory elements. Cytogenet. Genome Res. 124: 268–276.

Pezer, Z., and D. Ugarović, 2012  Satellite DNA-associated siRNAs as mediators of heat shock response in insects. RNA Biol. 9: 587–595.

Pezer, Z., J. Brjaković, I. Feliciello, and D. Ugarović, 2011  Transcription of satellite DNAs in Insects. Prog. Mol. Subcell. Biol. 51: 161–179.

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

Pons, J. B., E. Bravo, M. Petitpierre, D. Plohl, Ugarović et al., 2004  Complex structural feature 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., 2008  jModelTest: Phylogenetic model averaging. Mol. Biol. Evol. 25: 1253–1256.

Richards, S., R. A. Gibbs, G. M. Weinstock, S. J. Brown, R. Denell et al., 2008  The genome of the model beetle and pest Tribolium castaneum. Nature 452: 949–955.

Rossi, M. S., C. G. Pesce, O. A. Reig, A. R. Kornbiht, and J. Zorzópulos, 1993  Retroviral-like features in the monomer of the major satellite DNA from the South American rodents of the genus Ctenomys. DNA Seq. 3: 379–381.

Samuelson, L. C., R. S. Phyllips, and L. J. Swanberg, 1996  Amylase gene structures in primates: retropon insertions and promoter evolution. Mol. Biol. Evol. 13: 767–779.

Smith, P. G., 1976  Evolution of repeated sequences by unequal crossover. Science 191: 528–535.

Stuart, J. J., and G. Mocelin, 1995  Cytogenetics of chromosome rearrangements in Tribolium castaneum. Genome 38: 673–680.

Talavera, G., and J. Castresana, 2007  Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56: 564–577.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei et al., 2011  MEGAS: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28: 2731–2739.

Thompson-Stewart, D., G. H. Karpen, and A. C. Spradling, 1994  A transposable element can drive the concerted evolution of tandemly repetitive DNA. Proc. Natl. Acad. Sci. USA 91: 9042–9046.

Ting, C. N., M. P. Rosenberg, C. M. Snow, L. C. Samuelson, and M. H. Meisler, 1992  Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. Genes Dev. 6: 1457–1465.

Ugarović, D., 2005  Functional elements residing within satellite DNAs. EMBO Rep. 6: 1035–1039.

Ugarović, D., and M. Plohl, 2002  Variation in satellite DNA profiles—causes and effects. EMBO J. 21: 5955–5959.

Ugarović, D., M. Podnar, and M. Plohl, 1996  Satellite DNA of the red flour beetle Tribolium castaneum—comparative study of satellites from the genus Tribolium. Mol. Biol. Evol. 13: 1059–1066.

Wang, S., D. M. Lorenzen, R. W. Beeman, and S. J. Brown, 2008  Analysis of repetitive DNA distribution patterns in the Tribolium castaneum genome. Genome Biol. 9: R61.

Waterhouse, R. M., E. M. Zdobnov, F. Tegenfeldt, J. Li, and E. V. Kriventseva, 2011  OrthoDB: the hierarchical catalog of eukaryotic orthologs in 2011. Nucleic Acids Res. 39: D283–D288.

Zelentsova, E. S., R. P. Vashakidze, A. S. Krayev, and M. B. Evgen'ev, 1986  Dispersed repeats in Drosophila viridis: elements mobilized by interspecific hybridization. Chromosoma 93: 469–476.

Zuker, M., 2003  Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31: 3406–3415.

Communicating editor: J. A. Scott