Dynamic evolution of satellite DNAs drastically differentiates the genomes of *Tribolium* sibling species

Damira Veseljak[1], Evelin Despot-Slade[1], Marin Volarić[1], Lucija Horvat[1], Tanja Vojvoda Zeljko[1], Nevenka Meštrović[1], Brankica Mravinac[1]

[1]Ruđer Bošković Institute, Division of Molecular Biology, Bijenička cesta 54, HR-10000 Zagreb, Croatia

*Corresponding author:

Brankica Mravinac, Brankica.Mravinac@irb.hr

ABSTRACT

Tandemly repeated sequences, known as satellite DNAs (satDNAs), belong to the most rapidly evolving elements in eukaryotic genomes. Notwithstanding their plentitude and diverse, yet often inconsistent functions, the evolutionary dynamics of satDNAs and the way they model genomes are still poorly understood. In this study, we explored satDNA dynamics in satDNA-rich genomes by focusing on two insect species from the genus *Tribolium*, so closely related that they can produce sterile hybrids. We identified 135 satDNAs in the *Tribolium freemani* satellitome, accounting for 38.7% of the genome. By comparing the satellitome of *T. freemani* with that of *Tribolium castaneum*, we deciphered the origins of their predominant, species-specific major satellites, which evolved from different repetitive elements still present in both genomes. Importantly, we discovered that the centromeric regions of the two species share an orthologous organization, characterized by different major satDNAs and related minor satDNAs. The *T. freemani* male sex chromosome, which lacks the major satDNA but contains a minor-like satDNA, further heightened the question of which satDNA might be centromere-competent. By analyzing the long-range organization of centromeric regions, we discovered that both major and minor satDNA arrays exhibit a strong tendency toward macro-dyad symmetry, suggesting that the secondary structures in these regions may be more important than the primary sequence itself. We also found evidence that the centromeric satDNAs of *T. freemani* are present in extrachromosomal circular DNAs, which thus may be responsible for their expansion and homogenization between non-homologous chromosomes. Throughout the *T. freemani* genome, we identified numerous low-copy-number satDNAs that are orthologous between the siblings, some of which are associated with transposable elements, highlighting transposition as a mechanism of their spreading. These findings demonstrate that satDNAs with different genomic proportions and locations are subject to different DNA turnover mechanisms. While the dynamic evolution of satDNAs has undeniably played an important role in shaping and differentiating the genomes of *Tribolium* siblings, it remains to be determined whether the differences in their satDNA profiles are a cause or a consequence of speciation.

KEY WORDS

satellite DNA, satellitome, *Tribolium freemani*, *Tribolium castaneum*, sibling species, repetitive DNA evolution

RUNNING TITLE

Evolutionary dynamics of satellite DNAs in *Tribolium* sibling species

INTRODUCTION

Satellite DNAs (satDNAs) are highly repetitive, tandemly organized DNA sequences whose contiguous arrays make up extensive regions in many eukaryotic genomes. The repetitive organization has hampered their study in the past and was also the reason why these sequences have been the most poorly assembled parts of the sequenced genomes. They were initially considered "junk DNA" due to their abundance, albeit unclear function, but the accumulation of studies on numerous organisms began to change their unfavorable reputation over time. Since they frequently form heterochromatin blocks in the (peri)centromeric regions of many eukaryotes, their most recognized role is attributed to the organization of functional centromeres (Hartley and O'Neill 2019; Talbert and Henikoff 2022), which is consequently associated with proper chromosome segregation and genome stability (Flynn and Yamashita 2024). A growing number of recent studies reveal the presence of satDNAs also in euchromatic regions (Sproul et al. 2020; Cabral-de-Mello et al. 2023; Rico-Porras et al. 2024; Volarić et al. 2024), and their influence on the expression of neighboring genes has been observed (Feliciello, Akrap, and Ugarković 2015). Aberrant transcription of some pericentromeric satDNAs has been associated with a tumor-promoting function (Bersani et al. 2015; Iwata et al. 2024). The evolutionary significance of satDNAs is supported by evidence that these sequences may be involved in speciation (Ferree and Barbash 2009; Jagannathan and Yamashita 2021). However, despite various ascribed roles, no satDNA function has yet been found to be ubiquitous in all organisms studied.

SatDNAs are among the most rapidly evolving sequences in eukaryotic genomes, and can differ significantly in nucleotide sequence and copy number of their repeat units even between related species. The main theories explaining the evolution of satDNAs were proposed at a time when, due to methodological limitations, only the most abundant satellites were discovered in genomes. According to concerted evolution, the gradual accumulation of mutations between repeat units of a given satellite leads to greater interspecific repeat variability compared to intraspecific variability (Elder and Turner 1995). In some cases, drastic differences between various satDNAs that cannot be explained by the process of gradual degradation of the nucleotide sequence are explained by the satDNA library theory. The library theory implies the existence of a satellite collection of different satDNAs in an ancestral genome, of which certain satDNAs significantly change the number of copies by amplification or reduction, leading to different satellite profiles in descendant species (Fry and Salser 1977; Meštrović et al. 1998). Modern sequencing technologies and accompanying bioinformatics tools have improved the detectability of satDNA, so that often up to over a hundred satDNAs can be detected in the genomes of different organisms (Boštjančić et al. 2021; João Da Silva et al. 2023). The availability of an increasing number of high-throughput satellitome analyses contributes to the understanding of satDNA evolution and encourages

93 the revision of previous evolutionary concepts (Belyayev et al. 2020). For these purposes, comparative
94 analyses of related species are very informative (Camacho et al. 2022), whereby organisms with a high
95 proportion of satDNAs represent an additional challenge, but also a beneficial source of information.

96 The flour beetle *Tribolium freemani* belongs to the genus that includes some of the most important
97 pests of stored agricultural products (Sokoloff 1972). In addition to their economic importance, *Tribolium*
98 species are also an excellent platform for the study of satellite DNAs, which occupy up to 60% of their
99 genomes (Mravinac and Plohl 2010). The representative of the genus *Tribolium*, but also of the entire
100 order Coleoptera, is *Tribolium castaneum*. It was the first beetle whose genome was sequenced in 2008
101 (Richards et al. 2008), and the assembly has been refined and improved in the meantime (Herndon et al.
102 2020), with the latest assembly version TcasONT being significantly enhanced in the repetitive part of the
103 genome (Volarić et al. 2024). The satDNA studies have revealed that the *T. castaneum* genome contains
104 at least 57 different satellites (Ugarković, Podnar, et al. 1996; Feliciello, Akrap, Brajković, et al. 2015; Pavlek
105 et al. 2015; Gržan et al. 2023), of which the major satellite TCAST comprises 17% of the genome. The
106 closest congener to *T. castaneum* is *T. freemani*. The two species are so closely related that they can
107 hybridize but produce sterile offspring (Nakakita et al. 1981). We recently sequenced and assembled the
108 *T. freemani* genome and found that the two siblings are very similar in their coding sequence (Volarić et
109 al. 2022). Regarding satDNAs, it was discovered three decades ago that 31% of the *T. freemani* genome is
110 made of one satDNA (Juan et al. 1993), whose nucleotide sequence surprisingly shows absolutely no
111 similarity to the major satellite of *T. castaneum.* However, in both species the major satellites, although
112 remarkably different in sequence, occupy the (peri)centromeric regions (Juan et al. 1993; Ugarković,
113 Podnar, et al. 1996; Gržan et al. 2020).

114 Understanding the evolutionary dynamics by which satDNAs mold genomes is necessary for the
115 perception of genome versatility, and it may also be an essential step in unraveling the role of these
116 functionally controversial sequences. Considering the extremely variable nature of satDNAs and inspired
117 by the fact that *T. freemani* and *T. castaneum* produce infertile hybrid progeny, in this work we address
118 the question of how compatible the two species are in their satellite profiles. In an initial satellitome
119 analysis, we discovered 135 satDNAs in *T. freemani*. By revising the major satellite and analyzing its ultra-
120 long arrangement, we deciphered the organization of the centromeric regions and revealed which
121 satDNAs are involved in the structure of these architecturally and functionally crucial chromosomal parts.
122 We also discovered the origin of diverse major satDNAs of *T. freemani* and *T. castaneum*. Finally, by
123 exploring the orthologous satDNAs of the two species, we draw conclusions about the evolutionary
124 dynamics and the main molecular mechanisms that most apparently differentiated the siblings'
125 satellitomes and their genomes in general.

126 **RESULTS**

127 **1. Identification of *T. freemani* satDNA repeats**

128 To decipher the *T. freemani* satellitome, we sequenced its genome by Illumina sequencing. Randomly
129 subsampled sets of Illumina reads corresponding to different genome coverages were subjected to graph-
130 based clustering using the TAREAN program (**Suppl. Table S1**). Consensus sequences of the clusters
131 annotated as potential satDNA were mapped to the *T. freemani* genome assembly Tfree1.0 (Volarić et al.
132 2022), applying the criterion of ≥70% sequence similarity. The candidate sequences that were mapped in
133 ≥5 consecutive copies were declared satDNAs. In this way, 135 satDNAs were identified, one already
134 known from previous studies (Juan et al. 1993) and 134 new ones. The consensus sequences of the 135
135 satDNAs, named TfSat01-135, are listed in the **Suppl. Table S2**.
136 Of the 135 discovered satDNAs, 124 were found to be unrelated, while 11 showed partial similarity (56.7-
137 73.0%) in nucleotide sequence, suggesting a related origin. Based on the similarities, we classified the
138 related satDNAs into the four superfamilies, SF1-SF4 (**Suppl. Fig. S1, Suppl. Table S3**). The consensus
139 sequences of the new satDNAs were also searched against the NCBI GenBank nucleotide database, and
140 the only similarity of some *T. freemani* satDNAs was with previously described satellites of the sibling
141 species *T. castaneum* (**Suppl. Table S3**). When compared with repetitive sequences from the Repbase
142 collection, the 75 *T. freemani* satDNAs showed only segmental similarities with different mobile elements,
143 mainly DNA transposons (**Suppl. Table S3**).
144 Regarding repeat unit lengths, the consensus sequences of the 135 satDNAs range from 63 bp to 1106 bp
145 (**Suppl. Table S3**), with an evident preference for lengths between 140-180 bp (**Fig. 1A**). For five satDNAs,
146 we found that their monomers are higher-order repeats (HOR) based on two or three 90-188 bp long
147 subunits, sharing 63.0-82.2% pairwise similarity (**Suppl. Fig. S2**). Another feature that largely characterizes
148 the *T. freemani* satellitome is the biased DNA base composition, with 127 satDNAs having an A+T content
149 >60% (**Fig. 1A**, **Suppl. Table S3**).
150
151
152
153
154

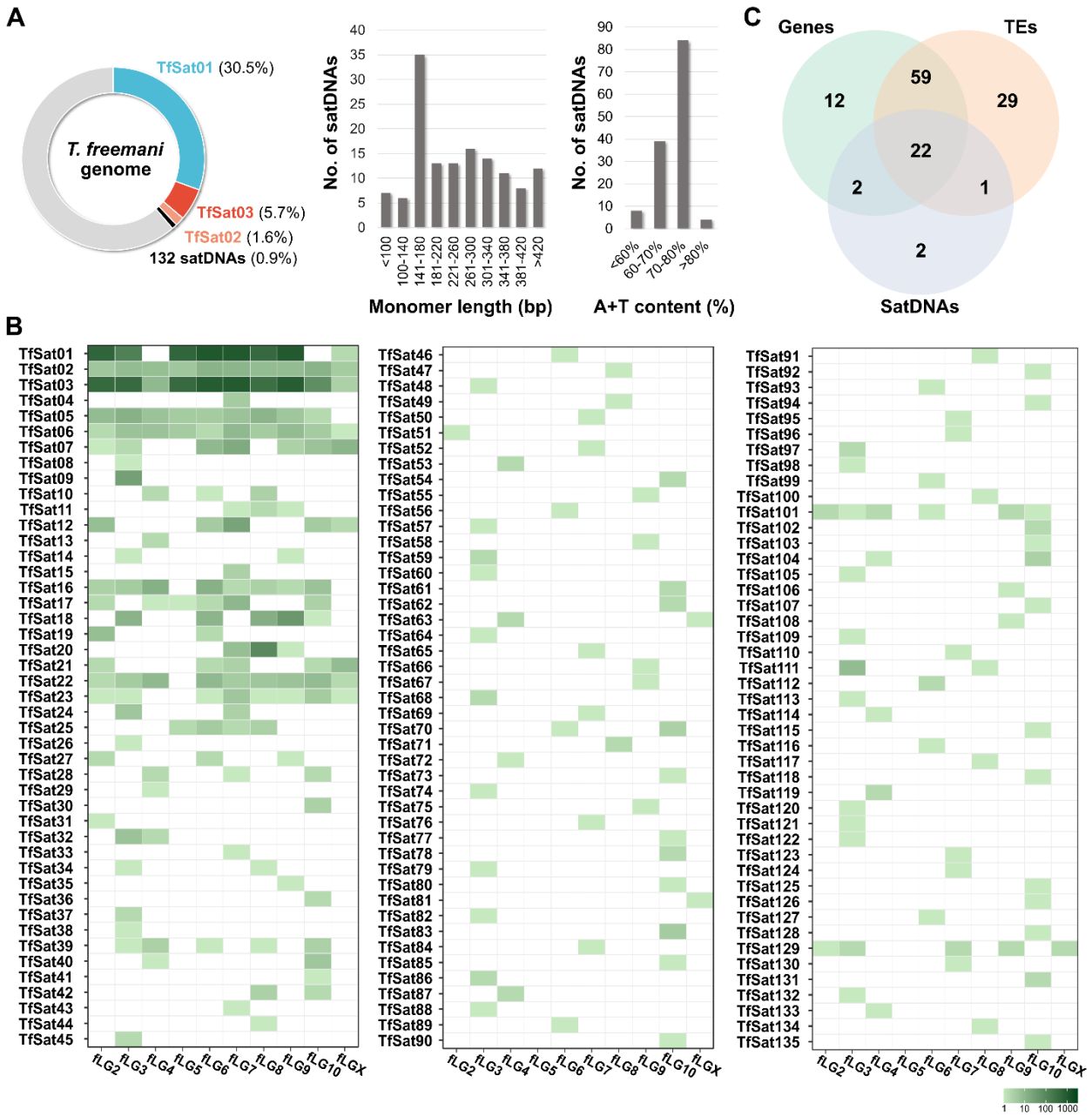


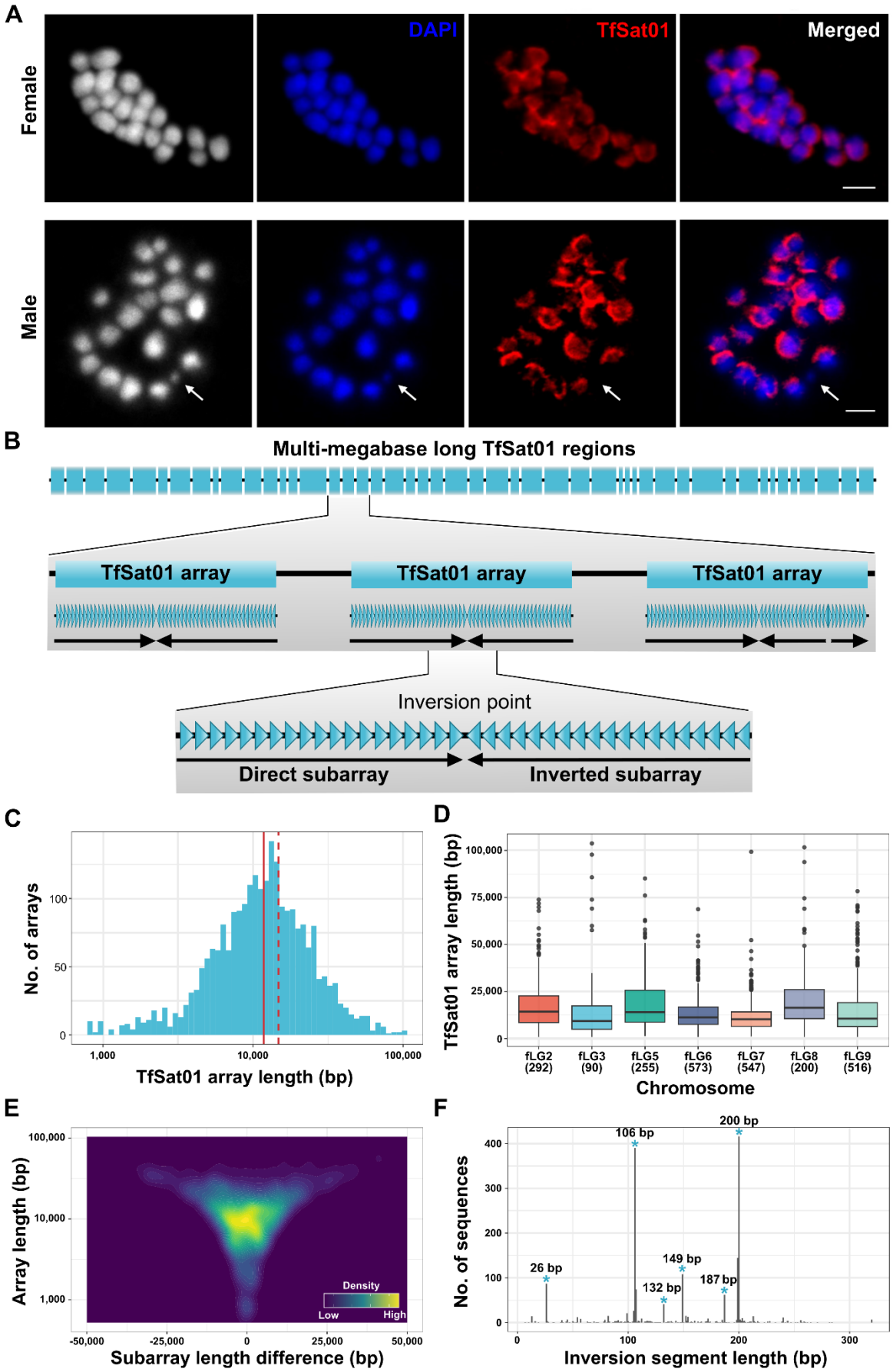
Figure 1. The overview of the *T. freemani* satellitome comprising 135 satDNAs. A) The genomic proportions of the three most abundant satDNAs and 132 low-copy-number satDNAs, and their distributions regarding the monomer length and A+T content of the consensus sequences. The genomic proportions are based on the estimates according to the TAREAN analyses, while the distributions of monomer lengths and A+T composition were derived from the data for the individual satDNAs listed in Supplementary Table S3. **B)** The presence of satDNA arrays with ≥5 consecutive copies mapped on the ten *T. freemani* chromosomes, fLG2-fLGX, in the genome assembly Tfree1.0. The number of annotated copies on chromosomes is color-coded according to the color scale. **C)** Venn diagram showing the number of the *T. freemani* satDNAs that have annotated genes, transposable elements (TEs) and/or satDNAs in the 10-kb regions flanking their arrays. The diagram is based on the data for each satDNA shown in Supplementary Table S6.

169 The satellitome with the 135 satDNAs comprises 38.7% of the *T. freemani* genome (**Fig. 1A**). However, the
170 genome is dominated by a single, extremely abundant satDNA, TfSat01, which makes up almost one third
171 of the genome sequence. In addition, the two satDNAs, TfSat02 and TfSat03, are moderately represented
172 (1.6% and 5.7%, respectively), while the remaining 132 satDNAs together account for only 0.9% of the
173 genome and can be considered low-copy-number satellites (**Fig. 1A**, **Suppl. Table S3**). Regarding the
174 distribution, the repeats of the three most abundant satDNAs were detected on almost all chromosomes,
175 whereas the low-copy-number satDNAs showed different distribution patterns (**Fig. 1B, Suppl. Fig. S3**).
176 The majority of them (101 satDNAs) were assigned only to one chromosome in the Tfree1.0 assembly, and
177 a smaller number of them were mapped to multiple chromosomes (**Fig. 1B**). To verify the chromosomal
178 distribution of the low-copy-number satellites obtained *in silico*, we experimentally analyzed six of them
179 with a genome proportion between 0.01% and 0.05%. We found that satDNAs with a genome proportion
180 higher than 0.02% had a broader chromosomal distribution (**Suppl. Fig. S3A-C**), while satDNAs with a lower
181 genome proportion (<0.02%) revealed a distinct signal only on one chromosome pair (**Suppl. Fig. S3D-E**),
182 which is generally consistent with the *in silico* analysis. Further, regarding the repetitiveness of monomeric
183 units, repeats of 71 satDNAs were found exclusively in arrays containing ≥5 consecutive monomers, while
184 for 64 satDNAs, in addition to their longer arrays, we also detected the shorter stretches with less than 5
185 consecutive copies (**Suppl. Table S4**).
186 To gain insight into the genomic environment of the identified satDNAs, we analyzed the 10 kb regions
187 upstream and downstream of the satDNA arrays for the presence of other satDNAs as well as transposable
188 elements (TEs) and genes annotated in the Tfree1.0 assembly. It turned out that 111 satDNAs have TEs,
189 95 have genes, and only 27 of them have other satellite sequences in their 10-kb flanking regions **(Fig. 1C,**
190 **Suppl. Table S5-S6, Suppl. Fig. S4**). The concurrent presence of TEs and genes was detected in the
191 surrounding regions of as many as 59 satDNAs, while only 22 satDNAs harbor TEs, genes, and satDNAs in
192 their 10-kb proximity. Even when present, the satDNAs generally did not stand out as significantly closer
193 neighbors. Namely, regarding the distance of a nearest annotated element to a satDNA array, similar
194 average distances of 4-5 kb were found for all three addressed sequence types (genes, satDNAs, TEs)
195 (**Suppl. Table S7**). Therefore, we conclude that the *T. freemani* satellitome is not exclusively associated
196 with gene-poor regions and that most low-copy-number satellites are not located in regions dominated
197 by tandemly repeated sequences.
198 The obtained general overview of the satellitome served as a starting point for detailed analyses of the
199 most prominent satDNAs in the *T. freemani* genome and the investigation of the evolutionary dynamics
200 and molecular mechanisms that shaped the satellitome.
201
202

203 **2. TfSat01, the main constituent of the *T. freemani* satellitome and centromere**

204 *Revision of the major satDNA TfSat01*

205 First, we reexamined the most plentiful satDNA, named TfSat01 in this study. The 166 bp long TfSat01
206 sequence corresponds to the previously described major satDNA, which was experimentally determined
207 to comprise 31% of the *T. freemani* genome (Juan et al. 1993), also supported by our TAREAN analyses
208 (**Suppl. Table S3**). However, our TfSat01 consensus, based on 228,060 monomers annotated in the
209 Tfree1.0 genome assembly, showed 5.4% nucleotide difference compared to the 30-year-old GenBank
210 entry X58539, the consensus derived from five cloned monomers (Juan et al. 1993) (**Suppl. Fig. S5A**). We
211 hold that the TfSat01 consensus better represents the *T. freemani* major satDNA (argued in detail in the
212 Material and Methods section), so we used it in all downstream analyses.
213 The major satDNA was previously reported to be located in centromeric chromosomal areas of *T. freemani*
214 (Juan et al. 1993). *T. freemani* has a typical coleopteran karyotype, 2n=20, with nine pairs of autosomes
215 and one pair of sex chromosomes (Shimeld 1989). The sex chromosome pair in females consists of two X
216 chromosomes, while in males there is an Xy_p parachute-like association of the X and a tiny y_p chromosome.
217 To verify TfSat01 presence on all chromosomes, we performed fluorescence *in situ* hybridization (FISH) on
218 female and male metaphase chromosome spreads (**Fig. 2A**). We indeed found TfSat01 present on all 20
219 chromosomes in females. In males, however, we detected TfSat01 signals on 18 autosomes and the X
220 chromosome, but surprisingly not on the male-specific y_p chromosome, revealing that the sex
221 chromosome y_p is deprived of TfSat01 satDNA (**Fig. 2A**).
222
223


A
Female
Male
DAPI
TfSat01
Merged
B
Multi-megabase long TfSat01 regions
TfSat01 array
TfSat01 array
TfSat01 array
Inversion point
Direct subarray
Inverted subarray
C
No. of arrays
TfSat01 array length (bp)
D
TfSat01 array length (bp)
fLG2 (292)
fLG3 (90)
fLG5 (255)
fLG6 (573)
fLG7 (547)
fLG8 (200)
fLG9 (516)
Chromosome
E
Array length (bp)
Density
Low
High
Subarray length difference (bp)
F
No. of sequences
26 bp
106 bp
132 bp
149 bp
187 bp
200 bp
Inversion segment length (bp)

224

225 **Figure 2**. **The organization of the major satDNA TfSat01 in the *T. freemani* genome. A)** Localization of TfSat01
226 satDNA on the *T. freemani* female (2n=18+XX) and male (2n=18+Xy_p) metaphase chromosomes determined by
227 fluorescence *in situ* hybridization. The first panels show the chromosomes in a black and white version to better
228 visualize the contours of the chromosomes, especially the male minute chromosome y_p. The chromosomes are
229 stained in DAPI (blue fluorescence), while TfSat01 signals are shown in red fluorescence. A white arrow points to the
230 y_p chromosome lacking the TfSat01 signal. The bar represents 3 µm. **B)** A schematic illustrating the long-range
231 organization of the multi-megabase long regions consisting of TfSat01 arrays. Within the TfSat01 array, the TfSat01
232 monomers (blue triangles) are repeatedly organized into subarrays that differ from each other by the orientation of
233 the monomers that form them. Thus, the term "TfSat01 array" refers to a continuous array of TfSat01 monomers,
234 regardless of the number of subarrays it contains. **C)** Distribution of the lengths of the TfSat01 arrays. The red solid
235 line represents the median (11.8 kb), while the mean (14.8 kb) is indicated by the dashed line. **D)** The box plot analysis
236 of TfSat01 array length distribution at different *T. freemani* chromosomes. The black line within a box represents the
237 median length, and the number of analyzed arrays is indicated in parentheses below the chromosome name. **E)**
238 Density plot of the differences between the length of the direct and inverted subarrays within the 1793 TfSat01
239 arrays. The x-axis shows the difference in length between directly and inversely oriented subarrays within an array.
240 The relative abundance of subarray length differences in the graph is indicated by a color gradient. Density plots for
241 arrays on individual chromosomes are presented in Supplementary Figure S7. **F)** The length distribution of the
242 inversion segments from the 1793 TfSat01 arrays. The blue dots in the graph indicate the six groups of the most
243 frequent inversion segments, whose consensus sequences and detailed alignments are shown in Supplementary
244 Figure S8 and Supplementary Data S1, respectively.
245

246

247 *Long-range organization of the major satDNA TfSat01*

248 The Tfree1.0 genome assembly lacks the assembled TfSat01 arrays on the three chromosomes (fLG4,
249 fLG10 and fLGX) (Volarić et al. 2022), but the multi-megabase long regions of the successfully assembled
250 TfSat01 arrays on the seven chromosomes allowed us to investigate the long-range organization of
251 TfSat01. First, we employed StainedGlass tool to visualize these multi-megabase regions spanning 3.8-9.7
252 Mb (**Suppl. Table S8).** Sequence identity heatmaps confirmed that they are built from arrays, which, when
253 compared in 10-kb segments, share predominantly >93% identity **(Suppl. Fig. S6)**. At higher resolution,
254 TfSat01 regions showed a more complex organization than a monotonous repetition of basic units (**Fig.**
255 **2B**). In terms of consecutiveness, the median length of all identified uninterrupted TfSat01 arrays is 11.8
256 kb (**Fig. 2C**), although some very long arrays of consecutive TfSat01 monomers were found, such as the
257 103.7 kb long array assembled on chromosome fLG3. Notwithstanding the fact that the number of mapped
258 TfSat01 arrays on individual chromosomes varies by up to 6.4-fold, the median length of continuous arrays
259 on all chromosomes is similar, ranging from 9.3 to 16.3 kb (**Fig. 2D**), thus suggesting that TfSat01 array
260 lengths are generally balanced across the genome.
261 Interestingly, the TfSat01 arrays have a specific substructure consisting of two or more subarrays that show
262 inverted orientation of the TfSat01 repeats (**Fig 2B**). TfSat01 arrays with two inversely oriented subarrays
263 are the most numerous (71.8% of all mapped arrays), while among the remaining arrays, arrays with an
264 even number of inversely oriented subarrays predominate over those with an odd number of inversely
265 oriented subarrays (**Suppl. Table S9**). The favored even number of subarrays within most TfSat01 arrays

266 may suggest a preference for dyad symmetry. An analysis of hypothetic hairpin structures in the 1793
267 arrays consisting of two subarrays showed that the hairpins would have some degree of asymmetry in
268 their stems due to the variable number of TfSat01 monomers composing the direct and inverted subarrays.
269 We found that the difference in length between direct and inverted subarrays is greater for longer arrays
270 (**Fig. 2E**), proving such correlation at all chromosomes (**Suppl. Fig. S7**).
271 We further investigated these 1793 TfSat01 arrays by looking specifically at the segments where the
272 subarrays change orientation. We discovered that the majority of these "inversion segments" fall into one
273 of six groups defined by a specific length (26 bp, 106 bp, 132 bp, 149 bp, 187 bp, and 200 bp; **Fig. 2F**, **Suppl.**
274 **Fig. S8A-B**). All six inversion segment types are based on an abruptly terminated TfSat01 monomer
275 followed by a reverse-oriented TfSat01 monomer truncated at a different nucleotide position (**Suppl Fig.**
276 **S8A-B**). Regardless of the different truncation positions, all six inversion segment types have two
277 characteristics in common: 1) there are no other extraneous intervening sequences between the reverse-
278 oriented truncated TfSat01 copies, and 2) the reverse-oriented truncated copies do not exhibit increased
279 nucleotide sequence variability (**Suppl. Fig. S8B**). Being reverse complemented, the truncated copies
280 extend the stems of potential hairpins almost to the tip, ending in 9–27 bp long loops (**Suppl. Fig. S8C**). It
281 is essential to note that the six inversion segment types are not chromosome-specific, as each of them was
282 detected on three to seven chromosomes (**Suppl. Fig. S8D).** Not only are the sequences of each inversion
283 type highly conserved with an average similarity of 95.0%-99.9%, but 100% identical copies of each
284 inversion type were identified on non-homologous chromosomes (**Suppl. Data S1**). From this we conclude
285 that the long-range organization of TfSat01 is hallmarked not only by the complex structure of inversely
286 oriented subarrays, but also by the well-preserved inversion sites throughout the genome.
287
288 **3. TfSat03 resides in the intercalary segments between the TfSat01 arrays**
289 The analysis of TfSat01 long-range organization further revealed that the intervals between TfSat01 arrays
290 are mostly about 4 kb long (**Fig. 3A-B**). We found that these intercalary segments mainly harbor repeats
291 of TfSat03, the second most abundant satDNA, whose 340 bp long monomers make up 5.7% of the *T.*
292 *freemani* genome (**Fig. 1A, Suppl. Table S3**). Although TfSat03 can form long continuous arrays, such as a
293 41 kb region on chromosome fLG3, it predominantly resides in the intercalary segments between TfSat01
294 arrays. These intercalary segments are structured as "cassettes" (**Fig. 3A**) containing a variable number of
295 highly conserved TfSat03 copies, but mostly between 6-8 (**Fig. 3C**). In addition to the conserved copies,
296 they also contain degenerate TfSat03 repeats that are truncated and/or have lower similarity to the
297 consensus (<80%). By their orientation, the conserved copies of TfSat03 together with the degenerate
298 copies form dyad symmetry within the intercalary segments (**Fig. 3A**). A ~110 bp long central region

299 (marked by the yellow rectangle in **Fig. 3A**), where the degenerate TfSat03 repeats change their
300 orientation, represents the midpoint of the dyad symmetry potentially resulting in a 13 bp loop (**Suppl.**
301 **Fig. S9**). Interestingly, the sequence of the central region is so degenerate that the partial similarity to
302 TfSat03 is barely recognizable. Despite the high degeneracy compared to TfSat03, the ~110 bp central
303 region is highly conserved between intercalary segments throughout the genome, showing an average
304 pairwise identity of 97.1% (**Suppl. Data S2**). To visualize the relationships and connections between central
305 regions from different intercalary segments, we used undirected graph networks generated from distance
306 matrices. The graph networks revealed that there are some small chromosome-specific clusters (**Suppl.**
307 **Fig. S10, Suppl. Data S3**), but the largest cluster with over two hundred 100% identical sequences of the
308 central region is formed from the intercalary segments distributed on different chromosomes (**Fig. 3D**). In
309 addition to the TfSat03 repeats, the intercalary segments enclose symmetrically arranged 200-300 bp long,
310 outward-facing stretches abundant in $(AAT)_n$ microsatellite (**Fig. 3A**). Overall, the conserved central regions
311 and the distal (AAT)n stretches appear to be the structural points of the intercalary segments, within which
312 the number of TfSat03 repeats oscillates. At the same time, the intercalary segments obviously represent
313 an integral part of the long-range organization of the major satDNA TfSat01.
314 In the Tfree1.0 assembly, we indeed detected *in silico* TfSat03 repeats on all chromosomes on
315 which TfSat01 arrays were annotated. As the assembly is still missing part of the genome, we considered
316 it important to experimentally verify the relationship between these two satDNAs *in situ*. The double-color
317 FISH confirmed the intermingled organization of TfSat01 and TfSat03, showing the signals of the two
318 satellites colocalized in broad regions of the 19 *T. freemani* chromosomes in males (**Fig. 3E**). Significantly,
319 no TfSat03 signals were detected on the y_p chromosome (**Fig. 3E**, marked by an arrow). The fact that the
320 y_p chromosome lacks the two most abundant and widespread *T. freemani* satDNAs opens the question of
321 y_p satellite profile.

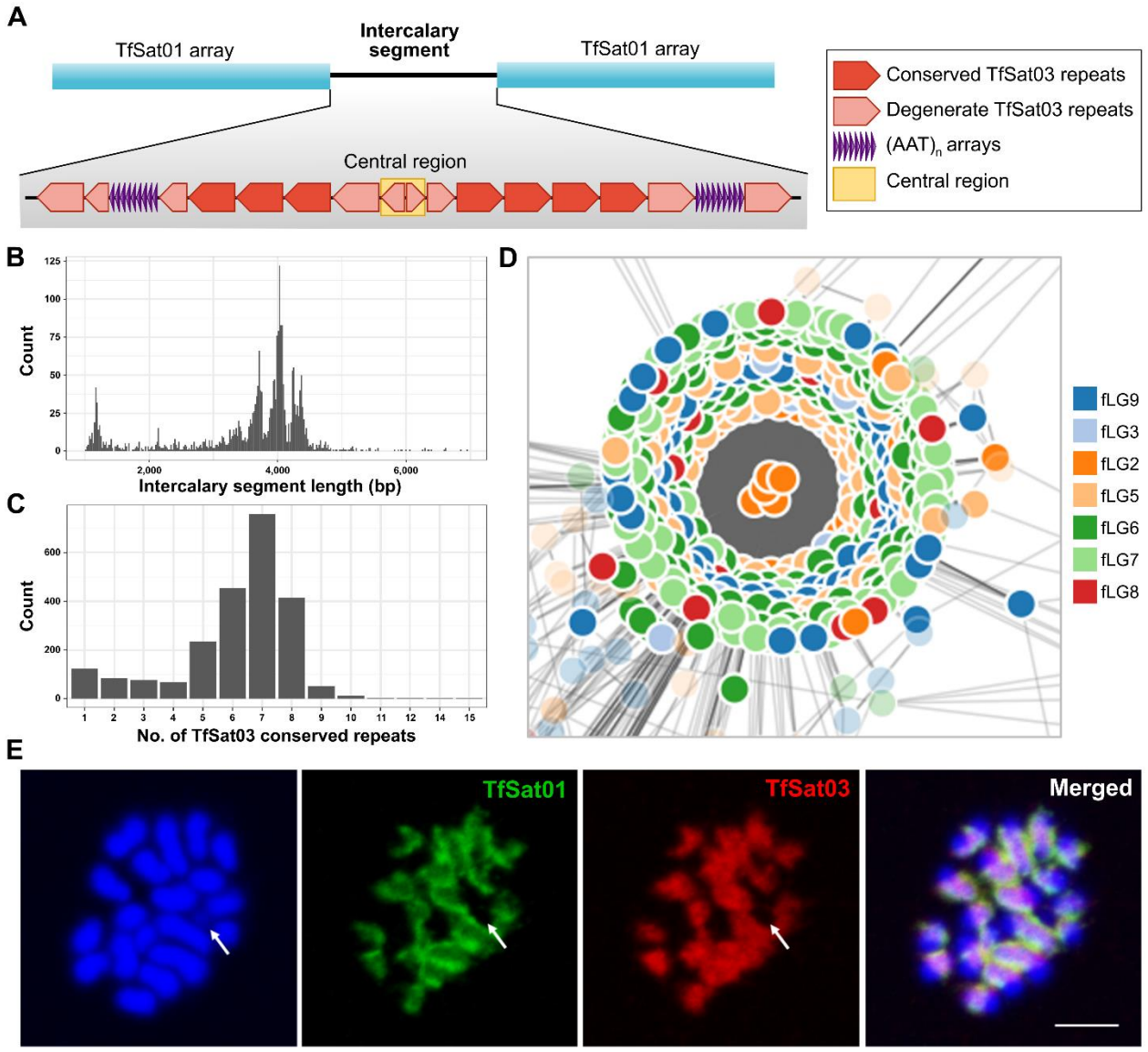
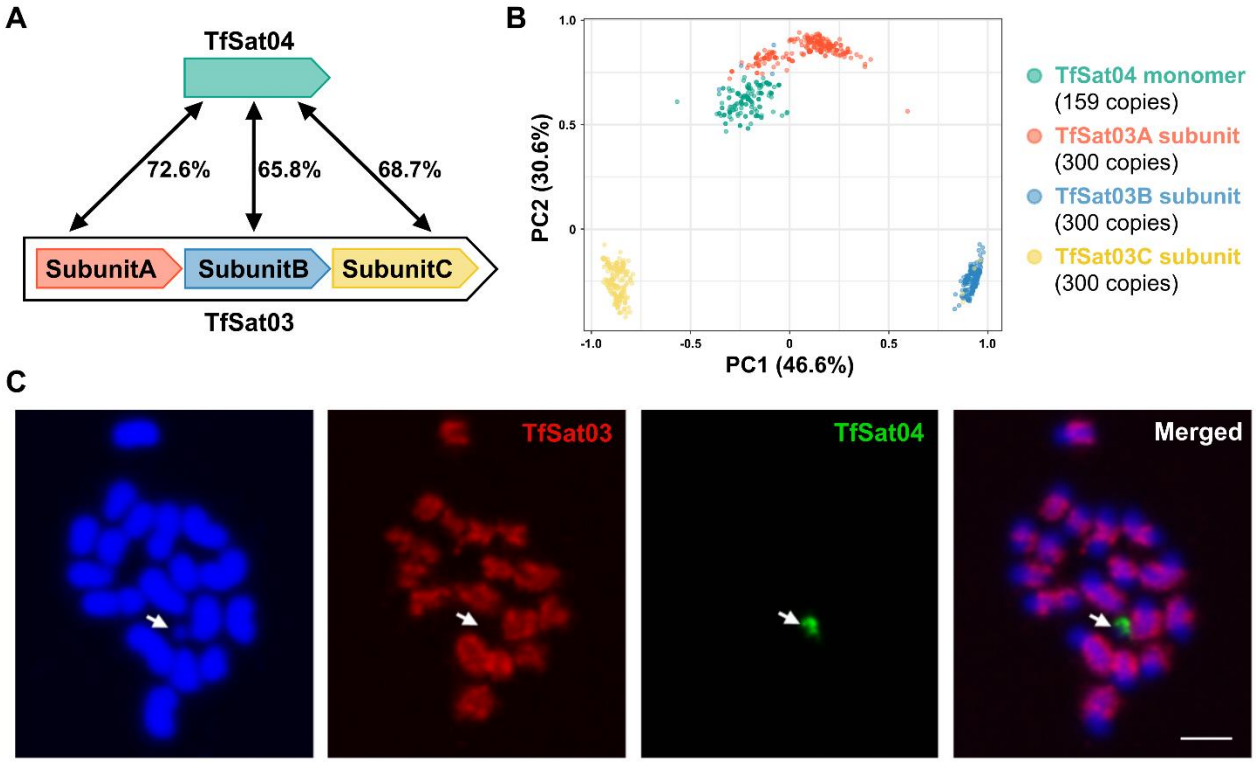


Figure 3. **Organization of TfSat03 satDNA repeats in the intercalary segments between TfSat01 arrays. A**) Schematic representation of the intercalary segments containing conserved and degenerate TfSat03 repeats and $(AAT)_n$ microsatellite arrays. The yellow rectangle marks the central region where the degenerate copies of TfSat03 change orientation. **B**) The distribution of intercalary segments lengths containing TfSat03 repeats. **C**) The number of conserved TfSat03 complete repeats per segment in the analyzed intercalary segments. **D**) Graph networks of central region sequences based on their sequence similarity relationship. The figure shows a detail of the graph network (the entire graph network can be found in Supplementary Figure S10) and highlights a cluster consisting of 100% identical sequences originating from different chromosomes. Each dot represents one sequence, and the color of the dot indicates the chromosome on which the sequence is located. The legend of the color-coded chromosomes is shown next to the graph. **E**) Co-localization of TfSat01 (green) and TfSat03 (red) satDNAs on the *T. freemani* male metaphase I chromosomes stained in DAPI (blue). An arrow points to the y_p chromosome lacking the TfSat01 and TfSat03 signals. The bar represents 3 µm.

339 **4. TfSat04, the male sex chromosome specific satDNA**

340 As mentioned, some *T. freemani* satDNAs show mutual similarities in nucleotide sequence. TfSat03 is one
341 of them, belonging to the SF1 superfamily, that also comprises satDNA TfSat04 (**Suppl. Table S3**). Contrary
342 to TfSat03, TfSat04 is a low-copy-number satellite that makes up only 0.07% of the genome. In addition to
343 the 80-fold difference in genome abundance, the two satDNAs also differ in the structure of their repeat
344 unit. The monomer TfSat03 is a 340 bp HOR whose three subunits A, B and C are 113-114 bp long and
345 share pairwise similarities of 65.8-73.0% (**Suppl. Fig. S2A**), while TfSat04 is based on a 112 bp long repeat
346 that corresponds to TfSat03 subunits sharing with them 65.8-72.6% similarity (**Fig. 4A, Suppl. Fig. S1A**).
347 The principal component analysis (PCA) of all 159 TfSat04 monomers mapped in the Tfree1.0 assembly
348 and 300 randomly subsampled TfSat03 subunits clearly separated the sequences into the four clusters
349 reflecting their provenience (**Fig. 4B**). According to the PCA, TfSat04 monomers are most closely related
350 to TfSat03 subunit A.
351
352



353
354
355 **Figure 4. Relationships between TfSat04 and TfSat03 satDNAs. A**) Schematic representation of the differences in
356 the structure of the basic repeat unit and the pairwise similarities between the TfSat04 monomer and the three
357 subunits of TfSat03. **B**) PCA clustering of 159 TfSat04 monomer copies and TfSat03 subunits A, B and C extracted from
358 the 300 randomly selected TfSat03 repeats. The monomers and subunits are represented by color-coded dots. The
359 color-coded legend is provided next to the PCA plot, and the proportions of variance for the first two principal
360 components, PC1 and PC2, are indicated on the axes in parentheses. **C**) Co-localization of TfSat03 (red) and TfSat04

361 (green) satDNAs on the *T. freemani* male metaphase I chromosomes stained in DAPI (blue). The arrow points to the
362 y_p chromosome, on which there is no TfSat03 signal, but a TfSat04 signal that is exclusively present on this
363 chromosome. The bar represents 3 µm.
364

365 Since the Tfree1.0 assembly contains only a single continuous array of TfSat04 annotated on chromosome
366 fLG7, by performing FISH we expected to detect it on one chromosome. Indeed, only one TfSat04 signal
367 was obtained on chromosome spreads, but to our surprise the signal was located on the y_p chromosome.
368 Double-color FISH confirmed that the positions of TfSat03 and TfSat04 are mutually exclusive, i.e. while
369 TfSat03 is localized at 19 chromosomes and not on y_p, TfSat04 is exclusively present on y_p (**Fig. 4C**). FISH
370 on female chromosome spreads, where no TfSat04 signal was detected, confirmed that TfSat04 satDNA is
371 specific for the male sex chromosome.
372 The experimental detection of TfSat04 on the y_p chromosome prompted us, as the authors of the Tfree1.0
373 genome assembly, to reexamine the original contigs that made up Tfree1.0. After revision (explained in
374 detail in Material and Methods section), we concluded that the TfSat04-containing contig, located at the
375 end of chromosome fLG7 in the Tfree1.0 assembly, belongs to the male chromosome y_p. Therefore,
376 following this finding, we will curate the Tfree1.0 assembly by upgrading it to a new version Tfree1.1, in
377 which the 2.2 Mb long end of chromosome fLG7 is separated and proposed as the y_p chromosome.
378
379
380 **5. Orthologous satDNAs between *T. freemani* and *T. castaneum***
381
382 To uncover the evolutionary trends in the satellitome of *T. freemani*, we included the orthologous satDNAs
383 of the most closely related species, *T. castaneum*, in the study.
384
385 *The origin of the centromeric major satDNAs of the sibling species*

386 The siblings *T. freemani* and *T. castaneum* differ by completely different, centromerically located major
387 satDNAs, TfSat01 and TCAST. Therefore, we first addressed the origin of these satellites.
388 By analyzing 228,060 TfSat01 monomer copies annotated in the Tfree1.0 assembly, we found that 93.3%
389 of them exhibit sequence similarity to the consensus greater than 90% (**Suppl. Fig. S5B**), proving that
390 TfSat01 is a very homogeneous satellite. Nevertheless, in the histogram of the similarity distribution, we
391 noted a distal peak corresponding to 82% similarity (indicated by black arrow in **Suppl. Fig. S5B**). Closer
392 examination revealed that these copies do not belong to tandemly organized TfSat01 repeats, but are
393 integrated parts of the 1106 bp long repeat units of *T. freemani* satDNA TfSat02 (**Fig. 5A, Suppl Fig. S11**).
394 Interestingly, TfSat02 repeats, which make up 1.6% of the genome (**Fig. 1A**, **Suppl. Table S3**), are located

395 outside the TfSat01 areas, often very far from centromeric regions (**Fig. 5B**). This raised the question of
396 the evolutionary relationship of the two satellites, i.e., which satDNA preceded the other: 1) was TfSat01
397 an ancestral satDNA whose monomeric unit invaded the TfSat02 precursor sequence, or 2) was a fragment
398 of the TfSat02 sequence excised and amplified into tandem arrays that gave rise to the present major
399 satDNA TfSat01. To find the answer, we searched the genome of the sibling species *T. castaneum* for
400 sequences that might be related to TfSat01 or TfSat02. We did not detect any tandemized TfSat01-like
401 copies in *T. castaneum*, but we discovered that TfSat02 is the ortholog of the previously described *T.*
402 *castaneum* satDNA TCsat15 (Gržan et al. 2023). The TfSat02 and TCsat15 repeats share a length of 1106
403 bp and a sequence similarity of 85.2% (**Suppl. Fig. S11**). Like TfSat02, the TCsat15 repeat harbors a 166 bp
404 segment corresponding to TfSat01 monomer (**Fig. 5A**, **Suppl. Fig. S11**), which proves that the
405 TfSat02/TCsat15 predecessor sequence contained the 166 bp TfSat01 precursor. From this we conclude
406 that the major satellite TfSat01 emerged from a segment of TfSat02. Comparison of the TfSat01 consensus
407 sequence with the corresponding 166 bp segments of TfSat02 and TCsat15 revealed that the TfSat02 and
408 TCsat15 segments were more similar to each other than to the TfSat01 consensus (**Fig. 5A**). In addition to
409 the simple consensus comparison, we performed a more comprehensive PCA, which included 10,000
410 randomly subsampled TfSat01 monomers, all TfSat02 166 bp segments annotated in the *T. freemani*
411 genome assembly Tfree1.0, and all TCsat15 166 bp segments annotated in the *T. castaneum* genome
412 assembly TcasONT. PCA strongly grouped the TfSat01 monomers into a separate, distant cluster (**Fig. 5C**).
413 In line with the fact that no tandemized TfSat01-like copies were mapped in *T. castaneum*, we conclude
414 that TfSat01 was not present as canonical satDNA in the common ancestor, but is a trait of the *T. freemani*
415 genome.
416

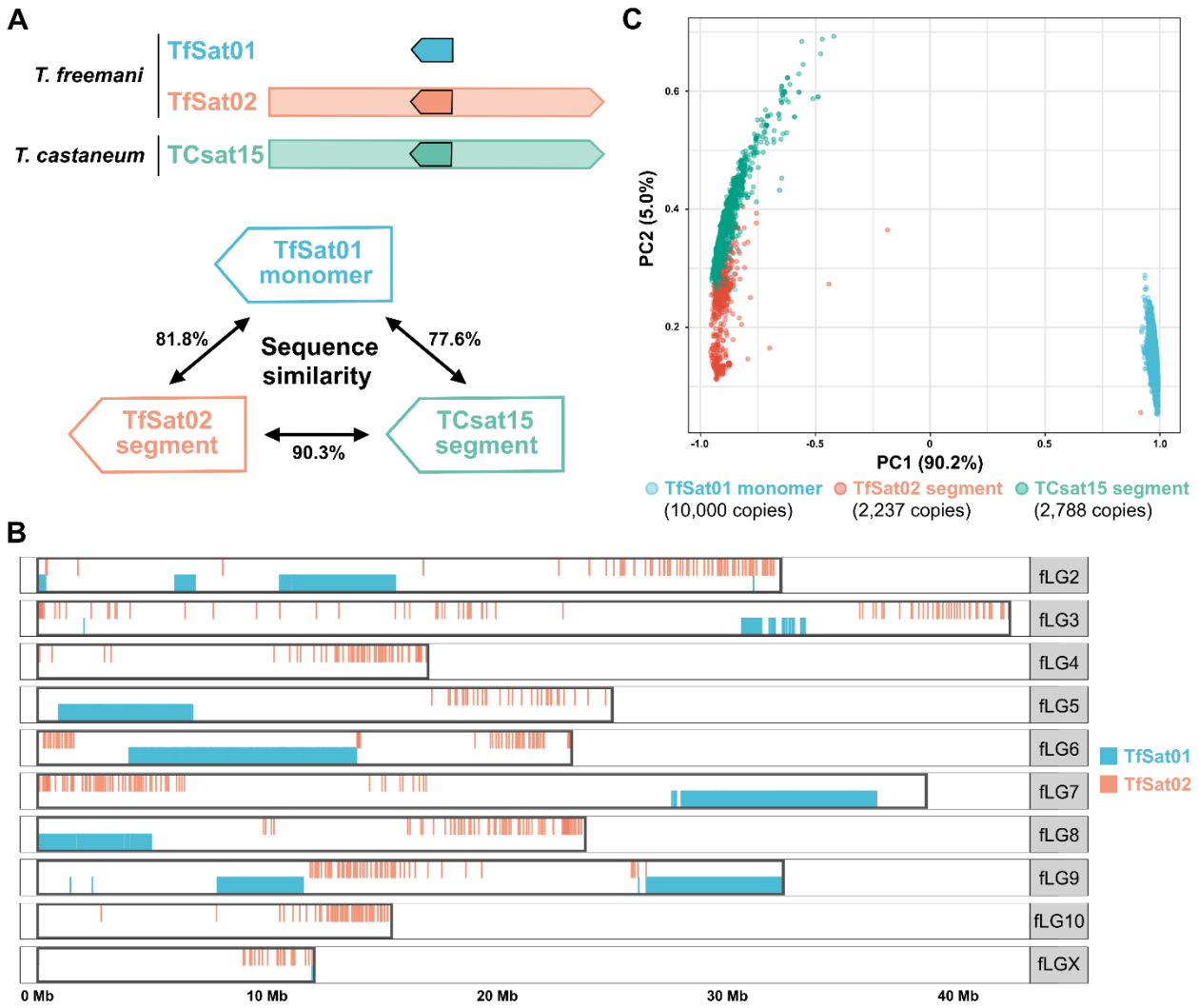


Figure 5. **Relationship between *T. freemani* satDNAs TfSat01 and TfSat02, and *T. castaneum* satDNA TCsat15. A)** Schematic representation of the TfSat01, TfSat02, and TCsat15 repeat units. The blue arrow represents the TfSat01 monomer, while the red and green arrows within the TfSat02 and TCsat15 units, respectively, indicate the position of the 166 bp segments that correspond to the TfSat01 monomer sequence. The percentages of nucleotide sequence similarity between the consensus sequences of TfSat01 monomers and the corresponding 166 bp segments of TfSat02 and TCsat15 are indicated. **B)** Distribution of satDNA TfSat01 and TfSat02 repeats mapped at the Tfree1.0 genome assembly along the *T. freemani* chromosomes fLG2-fLGX. **C)** PCA clustering of TfSat01 monomers and 166 bp segments from TfSat02 and TCsat15. The analysis included 10,000 randomly selected TfSat01 monomers and 2,237 copies of TfSat02 segments detected in the *T. freemani* genome assembly Tfree1.0, and 2,788 copies of TCsat15 segments annotated in the *T. castaneum* genome assembly TcasONT. Each dot represents a sequence of a single monomer/segment, and the color-coded legend is provided below the PCA plot. The proportion of variance for the first two principal components, PC1 and PC2, is shown on the axes in parentheses.

437 While searching for possible presence of the *T. castaneum* major satDNA TCAST in *T. freemani*, we did not
438 detect TCAST tandemized copies in the Tfree1.0 assembly, but we detected a TCAST transposon-like
439 element, previously identified in the *T. castaneum* genome (Brajković et al. 2012). Resembling to a DNA
440 transposon, the TCAST transposon-like element, being 1093 bp long, contains two segments
441 corresponding with 83% similarity to a TCAST monomer and its truncated version in reverse orientation,
442 partially overlapping with ~290 bp inverted termini (**Suppl. Fig. S12A**). In both genomes we found it
443 scattered on all chromosomes, with the exception of the *T. castaneum* X chromosome (**Suppl. Fig. S12B**).
444 When compared, the copies of the TCAST transposon-like elements annotated in *T. castaneum* and *T.*
445 *freemani* generally separate in the two species-specific groups (**Suppl. Fig. S12C**). Considering the 80%
446 sequence similarity and wide genomic distribution in both species, we hypothesize that the TCAST
447 transposon-like element was present in the ancestral genome, and served as a source from which the 360-
448 bp-based TCAST satDNA, now accounting for 17% of the *T. castaneum* genome, was propagated.
449
450 *The orthologous centromeric minor satDNAs*

451 TfSat03 is the second most abundant satellite in *T. freemani*, so next we searched for its ortholog in *T.*
452 *castaneum*. It turned out to be Cast7, one of the most variable *T. castaneum* satDNA with an estimated
453 genomic proportion of 0.2% (Pavlek et al. 2015; Volarić et al. 2024). It is based on ~109-114 bp repeat
454 units, which correspond to the TfSat03 subunits and show on average 67.2-79.3% similarity to them
455 (**Suppl. Fig. S13**). Even more fascinating than the nucleotide similarity between Cast7 and TfSat03 is their
456 orthologous organization characterized by intermingling with the major satDNAs and a propensity for dyad
457 symmetry (**Fig. 6A**). Namely, in *T. castaneum*, Cast7 arrays are found located mainly between the TCAST
458 arrays in the (peri)centromeric regions (Volarić et al. 2024). It is significant that Cast7 also shows changes
459 in the orientation of its copies within the arrays, and we discovered that the major satellite TCAST does
460 the same. The TCAST-Cast7 arrangement in *T. castaneum* is more irregular than the TfSat01-TfSat03
461 juxtaposing in *T. freemani*, but the positional and organizational orthology of the minor satellites Cast7
462 and TfSat03 is evident despite the fact that the major satellites, TCAST and TfSat01, are different (**Fig. 6A**).
463

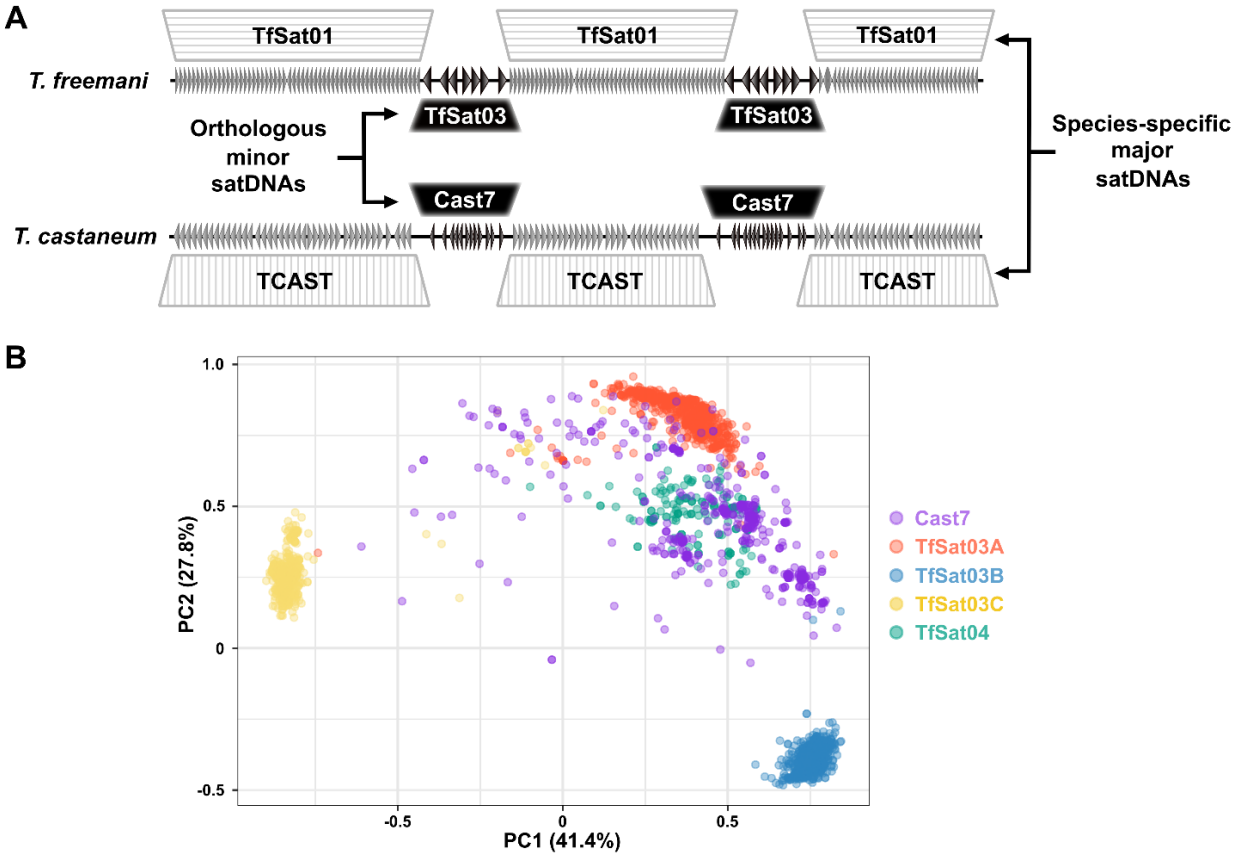


Figure 6. **The comparison of satDNAs in the centromeric regions of the sibling species *T. freemani* and *T. castaneum*. A)** A schematic illustrating the long-range organization of the species-specific major satDNAs TfSat01 and TCAST, intermingled with orthologous minor satDNAs TfSat03 and Cast7, respectively, in the centromeric regions of *T. freemani* and *T. castaneum*. **B)** PCA clustering of the repeats of the orthologous satDNAs Cast7, TfSat03 and TfSat04. The analysis included 1000 randomly selected Cast7 monomers from the TcasONT assembly, together with subunits A, B, and C from 1000 randomly selected TfSat03 repeats and all 159 TfSat04 monomers annotated in Tfree1.0 assembly. The sequences are represented by dots according to the color-coded legend provided next to the PCA plot.

Given the relatedness of the satellites TfSat03 and γ_p-specific TfSat04, we further investigated the relationship between Cast7, TfSat04 and TfSat03 subunits by analyzing randomly selected individual monomer/subunit sequences. According to PCA (**Fig. 6B**), Cast7 monomers intermingle between clusters of TfSat04 monomers and TfSat03 subunit A, confirming the previously observed highest similarity between their consensuses (**Suppl. Fig. S13**). The increased connection between Cast7-TfSat04-TfSat04_subunitA sequences could be an indication of ancestral sequence preservation, but potentially also a sign of a certain functionality.

<u>*The orthologous low-copy-number satDNAs*</u>

In the search for orthologous satellites, besides TfSat02, TfSat03 and TfSat04, we identified 11 other *T. freemani* satDNAs that have their orthologs among the previously characterized *T. castaneum* satellites

485 (**Suppl. Table S3)**. The alignments of the species consensuses show pairwise similarities from 63.8% to
486 94.8% (**Suppl. Fig. S14**). To gain a better insight into the relationships between monomeric copies of
487 orthologous satellites, we annotated orthologous repeats in Tfree1.0 and TcasONT and compared their
488 sequences (**Suppl. Data S4**). The resulting PCA plots for the orthologous satDNAs showed one of three
489 patterns: 1) species-specific clustering (TfSat08, TfSat11, TfSat15, TfSat25) (**Fig. 7A**), 2) extensive mixing of
490 repeats from both species (TfSat05, TfSat22, TfSat111) (**Fig. 7B**), 3) segregation of repeats from one species
491 in more than one cluster (TfSat07, TfSat27, TfSat42, TfSat76) (**Fig. 7C**).
492 We hypothesized that some of the remaining 120 *T. freemani* satDNAs might have orthologous copies that
493 are not defined as satDNA in *T. castaneum*, either due to the genuine lack of their tandem copies in *T.*
494 *castaneum* or due to the incompleteness of the repetitive DNA regions in the *T. castaneum* Tcas5.2
495 assembly used to define the *T. castaneum* satellitome (Gržan et al. 2023). To test this hypothesis, we
496 searched the *T. castaneum* TcasONT assembly using all *T. freemani* satDNAs as a query, and we indeed
497 found additional 45 satellites recognizing related copies in *T. castaneum*. Based on the individual copies'
498 alignments (**Suppl. Data S4**), the PCA (**Suppl. Data S5**) revealed that these 45 orthologs also follow one of
499 the three patterns mentioned above.
500
501

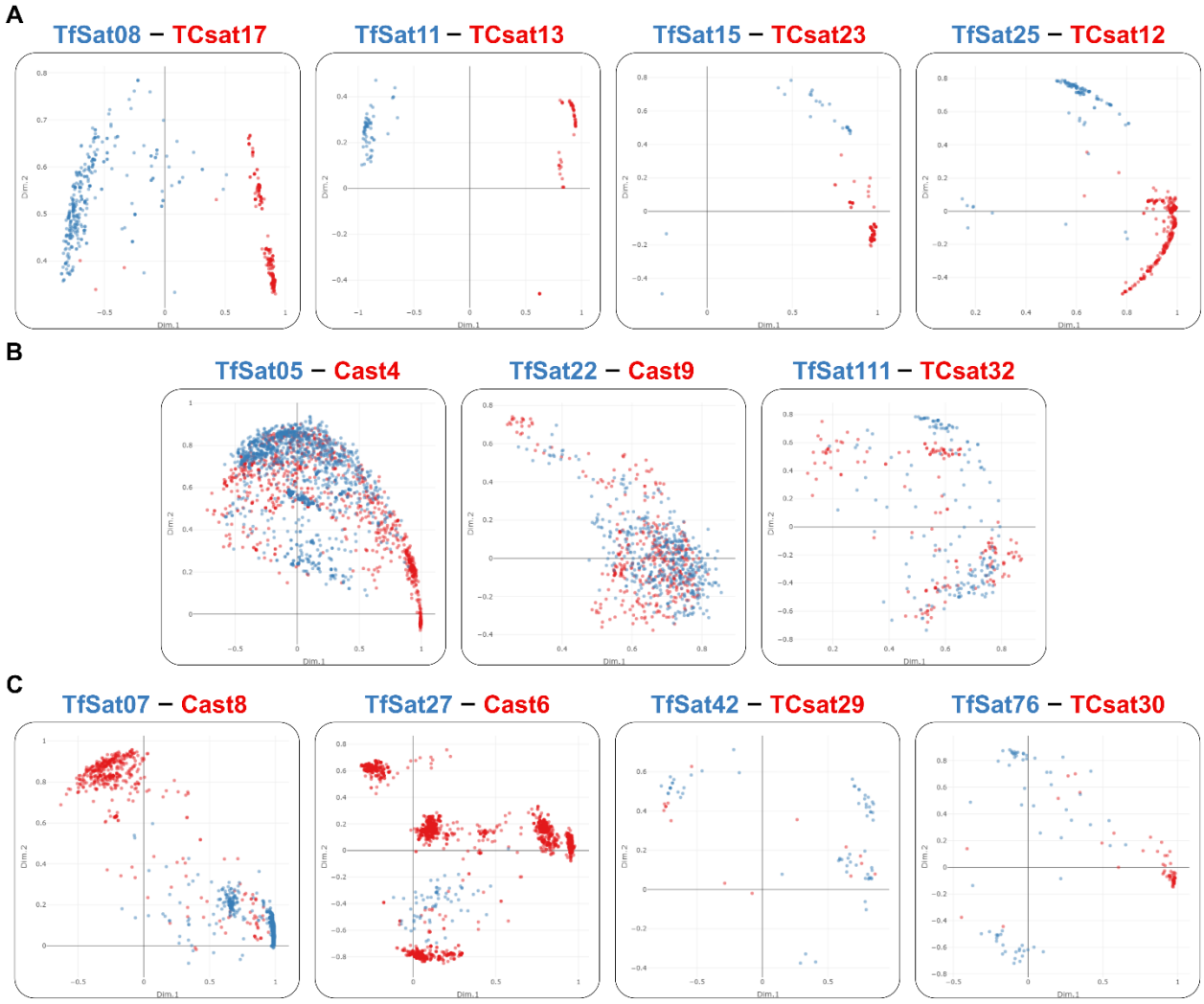


Figure 7. **PCA plots of the *T. freemani* satDNAs (TfSat) repeats and their orthologs among known *T. castaneum* satDNAs (Cast, TCsat)** showing: A) species-specific clustering, B) extensive interspecific mixing, C) segregation of repeats from one species in more than one cluster. The names of the orthologous satDNAs are indicated above each plot. The dots, colored according to the species of origin (*T. freemani* in blue, *T. castaneum* in red), represent monomer copies annotated in the *T. freemani* Tfree1.0 assembly and the *T. castaneum* TcasONT assembly. The PCA plots are based on the alignments shown in Supplementary Data S4. The interactive versions of the plots can be found in Supplementary Data S5.

6. Mechanisms of satDNA propagation

Being among the most variable sequences of eukaryotic genomes, satDNAs are subject to a very dynamic turnover. To infer possible mechanisms that have formed and probably actively remodel the satellitome of *T. freemani*, we first focused on the two most abundant satellites, TfSat01 and TfSat03. The intermingling of the two satellites on all chromosomes, with the exception of the sex chromosome y_p, indicates that the spread of TfSat01 and TfSat03 is closely linked. In addition, the inversion points within TfSat01 and TfSat03 arrays, distributed in a highly conserved form at non-homologous chromosomes,

520 suggest mechanisms that spread the two satellites together to different chromosomes. One of such
521 mechanisms could be the expansion via extrachromosomal circular DNA molecules (eccDNAs) that contain
522 the TfSat01/TfSat03 arrays and spread them by the reintegration into non-homologous chromosomes. To
523 test the presence of TfSat01 and TfSat03 repeats in the eccDNA fraction, we employed two-dimensional
524 (2D) agarose gel electrophoresis of the total genomic DNA. Subsequent Southern blot hybridizations with
525 TfSat01 and TfSat03 specific probes showed, as expected, TfSat01 and TfSat03 signals corresponding to
526 linear, chromosomal DNA, but also revealed their concomitance at the arc signals corresponding to
527 eccDNA, proving that these two satDNAs indeed occur in the form of extrachromosomal circular molecules
528 (**Fig. 8A**). Another way of spreading and exchanging these satellites between non-homologous
529 chromosomes could be achieved by 3D interactions between different chromosomes. Indeed, in meiotic
530 prophase I we occasionally observed the bouquet-like formations of non-homologous chromosomes
531 huddled together by (peri)centromeric heterochromatin. By double-color FISH, we confirmed that TfSat01
532 and TfSat03 signals colocalize at these close associations of the (peri)centromeric regions (**Fig. 8B**). We
533 conclude thus that the concomitance of the two satellites at multi-megabase long regions might be a
534 consequence of the joint action of both mechanisms, (peri)centromeric 3D interactions at meiotic
535 bouquet-like configurations and reintegration of eccDNAs.
536
537
538

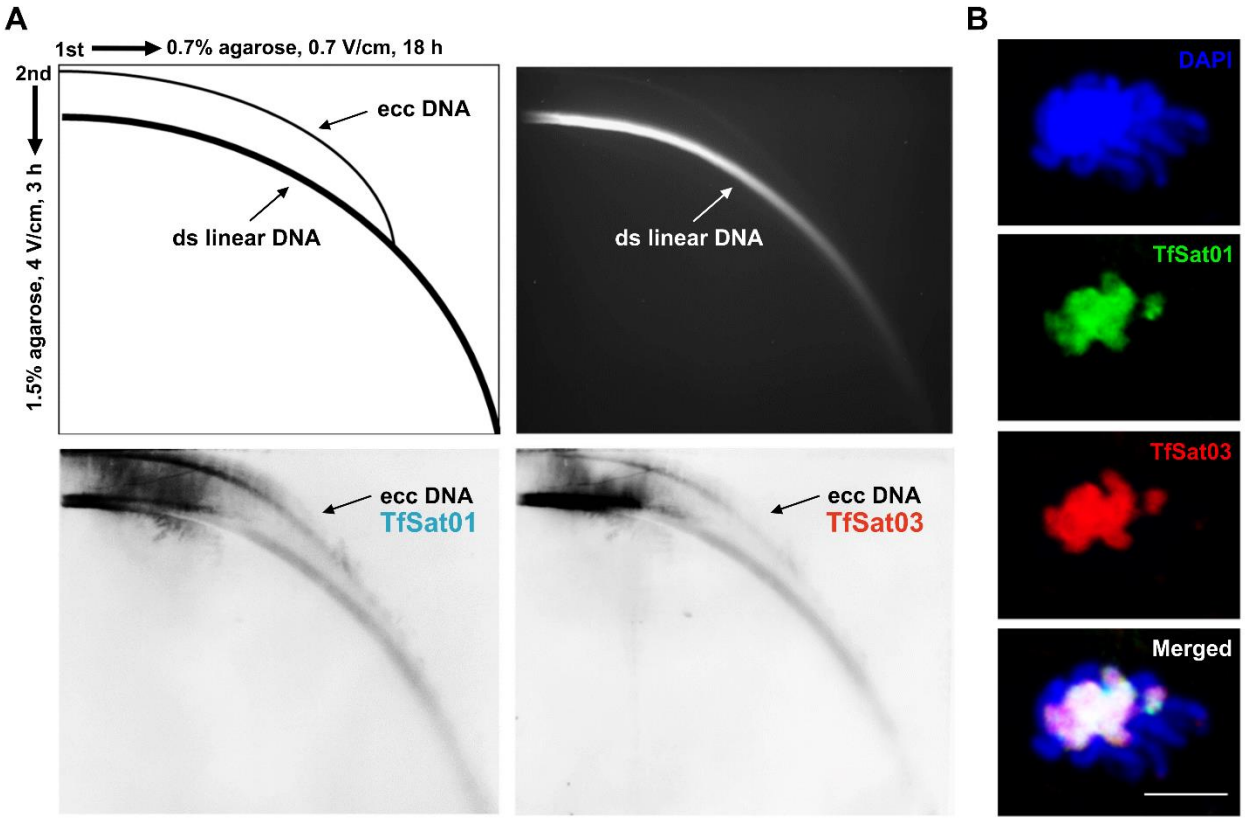


Figure 8. Experimental evidence of the presence of the TfSat01 and TfSat03 satDNA repeats in the extrachromosomal circular DNA (eccDNA) molecules and in prometaphase associations of non-homologous chromosomes. A) Two-dimensional gel electrophoresis analysis of the eccDNAs in *T. freemani*. The schematic at the top left illustrates 2D gel electrophoresis and the migration patterns of linear and circular DNA forms. The ethidium bromide stained gel (the upper right panel) with 2D-electrophoretically separated *T. freemani* genomic DNA was Southern-blotted and hybridized with the TfSat01 probe (lower left panel) and the TfSat03 probe (lower right panel). Both blots show the signals corresponding to eccDNAs (indicated by the arrows). **B**) Meiotic bouquet-like formations of *T. freemani* chromosomes with joint heterochromatic regions of non-homologous chromosomes. Chromosomes (stained with DAPI, top panel) were subjected to double-color FISH analysis using specific probes for TfSat01 (green) and TfSat03 (red). The bottom panel shows the overlap of the two satDNAs. The bar represents 5 µm.

Partial sequence similarities of certain low-copy-number satDNAs with TEs (**Suppl. Table S3**) suggest that transposons may be a source of some tandem repeats in *T. freemani* and that transposition may be a mechanism contributing to their propagation. In the satDNA-TE linkage analysis, we found the orthologous satDNAs to be particularly informative. The clearest examples of association with TEs come from satDNAs related to the Rehavkus superfamily of DNA transposons, characterized by long terminal inverted repeats and a variable number of subterminal tandem repeats. Here we highlight the three cases. The first of them refers to the satDNA TfSat15, which, like its ortholog in *T. castaneum*, TCsat23, forms short tandem arrays of several consecutive monomers in the inverted termini of the Rehavkus-1_TC transposon (**Suppl. Fig. 15A**). Analysis of the orthologs showed that the copies of TfSat15 and TCsat23 cluster into species-specific groups (**Fig. 7**), and this is also true for the rest of the transposon sequence (**Suppl. Fig. 15A**). Another example is the satellite TfSat23, which was found in the inverted termini of the Rehavkus-3_TC transposon.

562 The repeats of TfSat23 and its orthologous copies in *T. castaneum* also evolve according to the principles
563 of concerted evolution, as does the central part of the transposon sequence (**Suppl. Fig. 15B**). However,
564 Rehavkus-3_TC transposons in *T. freemani* mostly contain 6-8 TfSat23 repeats in their inverted termini,
565 while the number of satellite repeats in transposon copies in *T. castaneum* is generally larger, reaching up
566 to 50 consecutive monomers. The third and most illustrative example of how transposon-integrated
567 repeats can expand remarkably into long tandem arrays is TfSat25. In *T. freemani*, this repeat makes up 2-
568 12 consecutive copies in the inverted termini of the Rehavkus-1_TC-like elements. In *T. castaneum*, on the
569 other hand, its orthologous satellite TCsat12 forms very long stretches, the longest of which is 80.5 kb with
570 522 consecutive repeats on chromosome LG4 (**Suppl. Fig. 15C**). Considering that TfSat25 was found in the
571 *T. freemani* assembly exclusively in association with Rehavkus-like elements, we conclude that the intense
572 expansion of this repeat in *T. castaneum* occurred after the separation of the siblings. Although TfSat25
573 and TCsat12 still show a high similarity of 94.8% in nucleotide sequence (**Suppl. Fig. S14**), it is evident from
574 the clustering of their repeats (**Suppl. Fig. 15C**) that they have evolved in a species-specific direction after
575 the species split.
576 We also noted the effect of concerted evolution on satDNAs that are not linked to TEs. For example, the
577 orthologous satDNAs TfSat08 and TCsat17 were found in both species on chromosome (f)LG3 in an array
578 with a similar number of copies (339 in *T. freemani*, 290 in *T. castaneum*). Synteny analysis revealed that
579 the TfSat08/TCsat17 satellite block is embedded in the same environment in both species, and the 10 kb
580 regions upstream and downstream of the satellite block share an interspecific sequence similarity of 90.3%
581 and 92.1%, respectively (**Suppl. Fig. 16**). PCA analysis, however, showed a clear species-specific grouping
582 of individual TfSat08 and TCsat17 repeats (**Fig. 7A**). From this we conclude that the satDNA was already
583 present in the ancestral species in an array of about 40 kb in length, but after the separation of the species,
584 sequence conversion formed their species-specific variants.
585 Some satDNAs, such as TfSat06, show intense interchromosomal homogenization indicated by distribution
586 on all chromosomes of *T. freemani* without chromosome-specific clustering (**Suppl. Fig. S17A**).
587 Interestingly, TfSat06 arrays occur in an organizational form reminiscent of long, sometimes over 11 kb
588 long, inverted termini of transposon-like elements, which show the greatest partial similarity to Rehavkus-
589 1_TC sequence (**Suppl. Fig. S17B**), so we assume that transposition is responsible for the spreading of
590 TfSat06. Since we identified only one orthologous 20-repeat array in *T. castaneum*, we assume that the
591 expansion and interchromosomal homogenization of this satellite in *T. freemani* occurred after the
592 separation of the species. On the other hand, the orthologous satDNA pair TfSat22-Cast9 is evidence for a
593 satellite that probably began its homogenization between different chromosomes in the ancestral
594 genome. Namely, we saw a strong admixture of orthologous copies from the two species, but we also

595 detected a strong interchromosomal admixture when we analyzed repeats within a species (**Suppl. Fig.**
596 **S18**).
597 Regarding homogenization within a chromosome, for some low-copy satDNAs we noticed intraarray
598 homogenization. For example, the satDNA TfSat54 is present on chromosome fLG10 in two short arrays,
599 with the repeats from these arrays grouped in separate clusters (**Suppl. Fig. S19A**). Similarly, the satDNA
600 TfSat97 shows an array-specific grouping of repeats originating from two arrays on chromosome fLG3
601 (**Suppl. Fig. S19B**). In both cases, however, the arrays of these very low-copy-number satDNAs do not
602 exceed 20 tandem repeats, which raises the question of whether these are just satDNA "seeds" that will
603 spread or disappear in the future.
604
605
606 **DISSCUSSION**
607
608 At the genome level, advances in sequencing technologies over the past decade have had the
609 greatest impact on deciphering satDNAs, the most abundant but least functionally defined eukaryotic
610 sequences. In this study, we initially identified 135 satDNAs in the flour beetle *T. freemani* by using
611 assembly-free detection. Species with a larger number of detected satDNAs have been reported recently,
612 such as the frog *Proceratophrys boiei* having 226 satellites (João Da Silva et al. 2023) or the crayfish
613 *Pontastacus leptodactylus* with 258 satDNAs (Boštjančić et al. 2021), but their genomes are 10-60 times
614 larger than that of *T. freemani*. *T. freemani*, whose 135 satellites comprise 38.7% of the genome, certainly
615 represents one of the most satellite-rich genomes described so far. SatDNA cataloging provides general
616 information about the content of the satellitome, but analyses of the long-range organization and
617 structure of satDNAs arrays are essential to draw conclusions about the evolutionary dynamics of
618 satellites, which was the ultimate goal of this work. To achieve this, we took advantage of: 1) the high-
619 quality reference genome of *T. freemani* (Volarić et al. 2022) based on PacBio HiFi reads, and 2) the close
620 relationship of *T. freemani* with the model coleopteran *T. castaneum*, that also has a high-quality assembly
621 (Volarić et al. 2024) and a well-defined satellitome (Ugarković, Podnar, et al. 1996; Pavlek et al. 2015;
622 Gržan et al. 2023), which enabled us to discover orthologous satellites and study satDNA dynamics in these
623 satellite-rich genomes.
624
625 ***The evolutionary aspects of Tribolium satellitomes***
626 The origin of satDNA sequences, especially those that dominate genomes, is an intriguing but
627 rarely answered question, despite numerous satDNA studies from different organisms. Here, we revealed

628 the origin of the genome-prevailing and species-specific centromeric major satellites in the siblings *T.*
629 *freemani* and *T. castaneum*. We discovered that the *T. freemani* major satDNA TfSat01, which makes up
630 one third of the *T. freemani* genome, arose from a sequence segment of another satellite that is still
631 present in both genomes. Apart from the fact that the primordial satellite continues to exist in the current
632 genomes of the siblings, an equally intriguing question is why only 15% of its sequence length has evolved
633 into the new satellites. Proliferation of a dominant satDNA from part of a repetitive sequence is also
634 proved by the *T. castaneum* major centromeric satDNA TCAST. A DNA transposon-like element harboring
635 TCAST was found scattered in the genome of *T. castaneum* (Brajković et al. 2012), and we detected the
636 same DNA transposon-like element in the *T. freemani* genome. The previous work on TCAST-like elements
637 in *T. castaneum* also disclosed CR1-3_TCa retrotransposon, that contains a segment of 1.2 TCAST
638 monomer in its 5' UTR region (Brajković et al. 2012), and the authors pointed out the possibility that the
639 TCAST satellite derived from this retrotransposon. As we did not detect CR1-3_TCa retrotransposon copies
640 in *T. freemani* at all, we believe that TCAST originated from the DNA transposon, while CR1-3_TCa
641 retrotransposon in *T. castaneum* might captured part of TCAST sequence by insertion into the TCAST
642 satDNA array. The *T. freemani* and *T. castaneum* major satDNAs apparently evolved from different types
643 of repetitive sequences, TfSat01 from a satDNA and TCAST from a DNA transposon. The common link
644 between these source elements is their genome-wide distribution and presence on all autosomal
645 chromosomes. We conclude that the ancestral sequence TfSat02/TCsat15, which gave rise to satellite
646 TfSat01, must have been widely scattered in the ancestral genome, because in today's genomes both
647 widespread satellites, TfSat02 and TCsat15, often occur in shorter segments with one or a few repeats
648 (this work, (Gržan et al. 2023)), reminiscent of a dispersed element. We therefore hypothesize that the
649 widespread distribution of an element throughout the genome is, if not a prerequisite, at least a desirable
650 property that sequences with the potential to be the source of satDNAs exhibit. Under circumstances of
651 intense DNA turnover, for example in some stress situations, dispersed elements might have a greater
652 potential to distribute and exchange sequences with their counterparts from distant regions and/or non-
653 homologous chromosomes through mechanisms that drive satDNA turnover. Importantly, we found
654 neither tandemized copies of the *T. freemani* major satellite in the *T. castaneum* genome, nor tandemized
655 copies of the *T. castaneum* major satellite in *T. freemani*. Therefore, we conclude that the proliferation of
656 the major satDNAs in *T. freemani* and *T. castaneum* occurred simultaneously and possibly rapidly. The
657 striking amplification and propagation of a sequence from the ancestor's genome into a dominant satDNA
658 of the offspring genome could be explained by the satDNA library theory (Fry and Salser 1977), which has
659 been experimentally confirmed in different organisms (Meštrović et al. 1998; Silva et al. 2017). However,
660 the satDNA library hypothesis assumes the proliferation of the entire pre-existing repeat, whereas the

661 relationship between TfSat01 and TfSat02 in *T. freemani* proves that the propagation of only part of a
662 satellite monomer to a novel and highly abundant satDNA is possible, thus expanding the current
663 understanding of the satDNA library concept.

664 The low-copy-number satDNAs, whose number prevail in the *T. freemani* satellitome, show
665 different evolutionary statuses, from species-specific to orthologous between the two siblings. In both
666 groups there is a large number of satellites that show partial similarities with transposable elements (TEs).
667 Links between TEs and satDNAs have been observed and elaborated in the literature for some time
668 (Meštrović et al. 2015), and there is growing evidence that TEs are a prolific source of satellite sequences
669 (reviewed in Zattera and Bruschi 2022). In this respect, studies based on the long-range organization of
670 repetitive sequences are particularly revealing. For example, the work on the ultra-long Nanopore reads
671 of the legume *Lathyrus sativus* showed that nine of the 11 most abundant satDNAs occur predominantly
672 as short tandem arrays within the Ogre LTR-retrotransposons, and analysis of congeneric species
673 demonstrated that at least three satDNAs have evolved from the tandem subrepeats of the Ogre
674 retrotransposon (Vondrak et al. 2020). High-quality genome assemblies of the three species from the
675 *Drosophila virilis* group showed that much of the divergence in genome composition between the sister
676 species is due to Helitron TE-related tandem repeats (Flynn et al. 2024). Another TE-tandem repeat
677 association described in the *virilis* group involves the foldback DNA transposon *Tetris* and satDNA repeats
678 present in its large inverted termini, but also found in the genome tandemized in arrays of several kb (Dias
679 et al. 2014). In the context of our work, this link between the foldback transposons and the satDNAs is
680 particularly important. Namely, characterization of TE-satDNA associations in *T. freemani* highlighted
681 Rehavkus, the foldback DNA transposons that, like *Tetris*, carry tandem repeats at their inverted ends. It
682 has been shown that an increased number of tandem repeats in the inverted termini of foldback elements
683 affects their transposition by strengthening transposase binding and increasing excision frequency (Liu and
684 Wessler 2017), which may explain the benefit for foldback TEs to accommodate satellite repeats.
685 Interestingly, orthologous satDNAs associated with Rehavkus elements in *Tribolium* have shown that their
686 sequences between the two species change according to concerted evolution, as does the rest of the
687 transposon sequence. While some repeats can remain cocooned in the inverted termini of TEs for a long
688 time, some orthologs clearly demonstrated the amplification burst from transposon-residing repeats to
689 an 80 kb long satellite array. In view of the close connection between satDNAs and transposons, it is
690 reasonable to assume that TEs serve not only as a source of satellite sequences, but also as a means of
691 propagation. Transposition could be one of the efficient mechanisms by which satellite sequences spread
692 throughout the genome and thus also in euchromatin regions that were dogmatically thought to be
693 deprived of satDNAs. In *T. freemani*, we detected the majority of low-copy-number satDNAs in the vicinity

694 of genes and TEs, which is consistent with previous results for the low-copy-number satDNAs in *T.*
695 *castaneum* (Gržan et al. 2023). In *T. castaneum*, the extensive spread of even more abundant satDNAs in
696 gene-rich regions was revealed (Volarić et al. 2024), following the example of dynamic euchromatic
697 satDNAs previously described in *Drosophila melanogaster* and its three closest relatives (Sproul et al.
698 2020). A growing number of studies documenting the presence of satDNA arrays in euchromatic regions
699 testify that euchromatin is not inhospitable to tandemly repetitive sequences (Cabral-de-Mello et al. 2023;
700 Rico-Porras et al. 2024). However, functional analyses are required to clarify whether the spread of
701 satDNAs in euchromatin is function-driven or whether these arrays are just tolerated scattered repetitive
702 "seeds" with potential to expand.
703
704 ***SatDNA-rich centromeric regions in Tribolium***
705 As in many other eukaryotes, the centromeric regions of *Tribolium* abound in satellites. In *T.*
706 *castaneum*, satDNA-rich and unusually extended centromeres, so-called metapolycentromeres, comprise
707 almost half the length of individual chromosomes (Gržan et al. 2020), and the FISH results in this work
708 reveal a comparable situation in *T. freemani*. Such abundance brings into discussion the criteria that a
709 sequence must fulfill in order to ascend to a centromere-dominant satellite, but also the question of
710 whether a sequence or a structure is more important. The long-range organization of satDNAs in the *T.*
711 *freemani* (peri)centromeric regions pointed out dyad symmetries as the most conspicuous feature. Dyad
712 symmetries within satellite arrays have the potential to form secondary structures such as hairpins or
713 cruciforms, and enrichment with the predicted non-B-form DNA structures was found in the centromeric
714 regions of different animal and plant species (Kasinathan and Henikoff 2018; Liu et al. 2023). It has been
715 suggested that the role of such non-canonical DNA structures in centromeric regions may be to facilitate
716 the loading of centromere-specific nucleosome (Talbert and Henikoff 2022). Short inverted repeats (<10
717 bp), which could form non-B-form DNA structures, are present in the 166 bp long monomers of the *T.*
718 *freemani* major satDNA TfSat01 (Juan et al. 1993), as well as in the 360 bp long monomers of the *T.*
719 *castaneum* major satDNA TCAST (Ugarković, Podnar, et al. 1996). Furthermore, our fine-scale mapping
720 revealed that *T. freemani* centromeric regions are populated with frequent, several kb long TfSat01
721 subarrays, which exhibit striking macro-dyad symmetries, potentially ending in 9-27 bp loops that might
722 interact with centromere-specific proteins. In addition to these macro-dyad symmetries exhibited by
723 TfSat01 subarrays, the intercalary segments based on the minor satellite TfSat03 are organized in dyad
724 symmetry fashion too. The pronounced dyad symmetries also characterize the major satDNAs of other
725 *Tribolium* species, which have more complex monomeric units. In *Tribolium brevicornis*, the 1061 bp long
726 monomer of the major satDNA is composed of the two ~470 bp long inversely oriented subunits,

727 containing also numerous short direct and inverted repeats (Mravinac et al. 2005). A complex 704 bp long
728 monomeric unit based on dyad symmetry represents the satDNA MAD2 of the species *Tribolium madens*
729 (Ugarković, Durajlija, et al. 1996). In *Tribolium audax*, a related but much more complex monomer
730 characterizes the 1412 bp long repeat of the satDNA TAUD2, being based on two ~700 bp long inverted
731 subrepeats (Mravinac and Plohl 2010). All these *Tribolium* major satellites are present in wide centromeric
732 regions of all chromosomes. Therefore, we conclude that dyad symmetry may indeed be a structural
733 imperative and that secondary structures in *Tribolium* centromeres may be more important than the
734 primary structure. The superior importance of secondary structures could thus explain the lack of similarity
735 in nucleotide sequences between the major centromeric satDNAs of congeneric species.

736 Inversions of satDNA subarrays in (peri)centromeric regions are not unknown. In the recent study
737 of satDNA and chromatin organization in mouse centromeres, a low percentage of direction changes in
738 pericentromeric and centromeric satDNAs' arrays were detected, and the authors commented that they
739 might result from the inversion events (Packiaraj and Thakur 2024). In the *T. freemani* centromeres,
740 however, the direction switches are too pervasive to be explained by inversion events alone. We believe
741 that the frequent macro-dyad symmetries in the *T. freemani* centromeres not only represent a possible
742 structural preference, but may also indicate the mechanisms that promote the proliferation of
743 centromeric satellites. One of these mechanisms could be via eccDNA molecules. SatDNA-derived eccDNAs
744 have been found in a variety of eukaryotic species (Cohen et al. 2003; Cohen et al. 2006; Navrátilová et al.
745 2008; Huang et al. 2021; Liu et al. 2023), including *T. castaneum* (Volarić et al. 2024). In *T. freemani*, we
746 confirmed the presence of centromeric major and minor satDNAs in the eccDNA fraction, so we suggest
747 that the eccDNAs propagate adjacent arrays of the two satellites, promoting the dyad symmetries present
748 therein. Our conclusion is additionally supported by preserved direction changes in the major satDNA
749 arrays and in the minor satDNA-based intercalary segments, highly conserved in sequence even among
750 non-homologous chromosomes. In addition to eccDNA-mediated spread, 3D spatial interactions in the
751 nucleus among distant satellite arrays have been proposed as a mechanism that enables spread of satDNAs
752 between remote loci (Sproul et al. 2020). In *T. freemani*, we detected satDNA-rich (peri)centromeric
753 heterochromatin clustering in the meiotic prophase, which was also observed in other *Tribolium* species
754 (Žinić et al. 2000; Mravinac and Plohl 2010). Non-homologous centromere pairing during early meiotic
755 prophase I has been reported in multiple organisms (Kurdzo and Dawson 2015), including associations of
756 satDNA-rich centromeric regions of the X chromosome and different autosomes in mouse (Spangenberg
757 et al. 2021). In the study of murine telocentric chromosomes, which revealed high sequence identity
758 between non-homologous chromosomes, it was concluded that this reflects a mechanism of frequent
759 recombinational exchange between non-homologous chromosomes possibly facilitated by close

760 associations during meiotic prophase (Kalitsis et al. 2006). The clustering of the non-homologous
761 centromeres of *T. freemani* in early meiosis and the conservation of direction switches between satDNAs
762 arrays at non-homologous chromosomes are in line with this interpretation.
763
764 ***Potential functional implications of dynamic satDNA evolution***
765 One of the important findings of this work is that *T. freemani* and *T. castaneum* despite completely
766 different major satDNAs share an orthologous organization in their centromeres. In addition to dyad
767 symmetries, in both species the arrays of species-specific major satDNAs intermingle with arrays of
768 orthologous minor satDNAs. The fact that the siblings' centromeres are remarkably different in the major
769 satDNAs but similar in the less copious satellites contradicts the intuitive assumption that the most
770 abundant satDNAs are the key components of the functional centromeres. At this point, we are inclined
771 to hypothesize that the orthologous minor satDNAs (TfSat03 and TfSat04 from *T. freemani*, and Cast7 from
772 *T. castaneum*) could be centromere-competent sequences. An analogy can be drawn with the centromeres
773 of *Mus musculus*, in which the more abundant major satellite forms the pericentromeric heterochromatin,
774 while the less abundant minor satellite is centromeric (Joseph et al. 1989). Interestingly, the mouse Y
775 centromere is based on a highly diverged minor satellite-like sequence (Pertile et al. 2009). Similarly, the
776 *T. freemani* sex chromosome y_p does not harbor the major satDNA, but comprises y_p-specific satDNA
777 TfSat04, from the superfamily of minor satDNAs. It could be that the y_p-specific variant of minor satellite
778 is developed and maintained by intrachromosomal homogenization due to suppressed exchange
779 interaction with other chromosomes, also suggested by the absence of TfSat01/TfSat03 arrays on the y_p
780 chromosome.
781 Functional centromeres are usually determined epigenetically by the presence of centromere-
782 specific variants of histone H3, CENH3 (Mellone and Fachinetti 2021). By analyzing the centromeres of *T.*
783 *castaneum*, we found that the regions of major satDNA TCAST encompass the CENH3 metapolycentric
784 domains, but also extend beyond the CENH3 distribution (Gržan et al. 2020). CENH3-ChIP-seq analysis in
785 the same study showed that CENH3 chromatin is established on a variety of repetitive DNAs. However,
786 Cast7 was not recognized as a CENH3-associated sequence, but it is possible that the marked variability of
787 Cast7 hindered its recognition among CENH3-enriched clusters. Our preliminary testing of *T. castaneum*
788 CENH3 antibodies in *T. freemani* indicates that *T. freemani* has its own species-specific CENH3, consistent
789 with the centromere paradox hypothesis, which justifies the rapid evolution of centromeric DNA and
790 protein components even among closely related species (Henikoff et al. 2001). Detailed studies on
791 centromere-specific proteins, such as CENH3 histones, should help to decipher which of the *T. freemani*
792 satDNAs, major or minor, are actually involved in centromere function.

793 *T. freemani* and *T. castaneum* separated 14 Mya ago and represent the youngest species within
794 the >100 Mya old genus (Hinton 1948; Ramesh et al. 2021). In contrast to the cosmopolitan *T. castaneum*,
795 data on the geographical distribution of *T. freemani* are quite sparse, being limited to Kashmir, the region
796 of its first record (Hinton 1948). Although it is not known whether the two siblings coexist in the same
797 habitats, they interbreed under laboratory conditions and produce infertile F_1 hybrids (Nakakita et al.
798 1981), indicating postmating reproductive isolation (Wade and Johnson 1994). Considering that the
799 satDNAs are by far the most abundant sequences in their genomes and thus represent the largest genomic
800 difference between the two species, it is tempting to speculate whether they might play a role in
801 reproductive isolation. In the recent study on the two congeneric catfish species, genome-wide satDNA
802 divergence between the parental species, rather than chromosome number discrepancy, was indicated as
803 contributing to the sterility of the hybrids (Lisachov et al. 2024). Similarly, satDNAs that differ significantly
804 between sugar beet and wild beet genomes are thought to conduce to barriers in interspecific beet
805 breeding (Schmidt et al. 2024). In hybrids of closely related *Drosophila* species, differences in satDNA
806 composition between parental species have been shown to impede the proper clustering of
807 pericentromeric satDNAs into chromocenters in interphase nuclei, thereby promoting hybrid
808 incompatibility (Jagannathan and Yamashita 2021). To determine whether satDNAs discrepancies underlie
809 the reproductive isolation between *T. freemani* and *T. castaneum*, functional studies similar to those in
810 *Drosophila* are required, and the hybrids will play a key role in answering this question.
811
812 In conclusion, the comprehensive study of the *T. freemani* satellitome and the comparison with *T.*
813 *castaneum* revealed a dynamic evolution of satDNAs that yielded the greatest differences between the
814 siblings' genomes. Dissimilar levels of conservation, organization, chromosomal locations and abundances
815 point to different evolutionary dynamics and mechanisms of propagation to which individual satellites
816 within a genome are subject. Despite the complex and plentiful satellitomes, we deciphered the origin of
817 completely different major satellites together with the orthologous organization of centromeric regions,
818 which show macro-dyad symmetries and related minor satDNAs as a commonality between the siblings.
819 This work serves as a good reference point for satellitome analyses of other *Tribolium species*, which
820 should further improve the understanding of satDNA evolution in satDNA-rich genomes. In addition, this
821 study provides a fundamental basis for future work that will address the role of *Tribolium* most prominent
822 satDNAs in the context of functionality and speciation, in which *T. freemani*-*T. castaneum* hybrids will be
823 of great assist. We hold that satellite-abundant non-model organisms such as *Tribolium* can be very useful
824 in tackling questions about functional and evolutionary implications of satDNA behaviour, and our further
825 research will go in this direction.

826 **MATERIAL & METHODS**

827

828 **Insect material**

829 The *T. freemani* beetles, originally obtained as a starter culture from the USDA Agricultural Research
830 Service (Manhattan, KS, USA) in 2015, were propagated in the laboratory in whole wheat flour in a
831 darkened incubator at 27°C and 50-70% relative humidity. The beetles were subcultured every 4-6 weeks.

832

833 **DNA isolation and whole-genome sequencing**

834 Total genomic DNA was extracted from 50 mg of snap-frozen adult *T. freemani* beetles (10 pooled male
835 and female individuals) using the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany). The isolated
836 genomic DNA was quantified with Qubit 2.0 DNA HS Assay (ThermoFisher, Massachusetts, USA) and
837 quality was assessed by Tapestation genomic DNA Assay (Agilent Technologies, California, USA). DNA was
838 sent to the sequencing service provider Admera Health (South Plainfield, USA), where library preparation
839 was done using KAPA Hyper Prep Kit (Roche, Switzerland). Illumina® 8-nt unique dual-indices were used
840 to mitigate index hopping. Whole-genome sequencing (WGS) performed on an Illumina NovaSeq X Plus
841 10B platform (Illumina, California, USA) yielded 2x35,516,583 paired-end reads (2x151 nt). The resulting
842 10.7 Gb of sequenced data corresponds to approximately 35-fold coverage of the *T. freemani* genome. We
843 have deposited the raw sequencing data in the NCBI Sequence Read Archive (SRA) database under the
844 BioProject study accession number PRJNA1179347.

845

846 **SatDNA mining using graph-based clustering**

847 The identification of *T. freemani* satDNAs from the unassembled short Illumina reads was performed by
848 graph-based sequence clustering with the TAREAN pipeline (Novák et al. 2017). The raw reads were
849 quality-checked and preprocessed using the RepeatExplorer2 tools on the Galaxy web server
850 (https://repeatexplorer-elixir.cerit-sc.cz/galaxy/). The interlaced reads were randomly subsampled to
851 reduce the size of the input dataset in order to achieve a low genome coverage required for TAREAN
852 analyses. We used six randomly subsampled sets with 175,000 to 1,500,000 input reads, corresponding to
853 genome coverages of 0.02–0.7x (**Suppl. Table S1**). Since the major satDNA TfSat01 accounts for 31% of the
854 *T. freemani* genome, its high proportion considerably limits the number of reads that can be processed
855 with the TAREAN tool, as determined in the first test analysis (**Suppl. Table S1,** T1 analysis). For this reason,
856 in the subsequent analyses we used the input sets from which we excluded the TfSat01-containing reads
857 (**Suppl. Table S1,** analyses T2-T6). A custom database was created containing the consensus sequences of
858 the high and low putative satDNAs from all six TAREAN analyses, and blast search of the custom database

859 was performed to exclude duplicates. The consensus sequences of the TAREAN satDNA candidates were
860 mapped on the *T. freemani* genome assembly Tfree1.0, as described in the "Repeat annotations"
861 paragraph below. If a candidate was found tandemly repeated in an array of at least five consecutive
862 monomers, it was declared as satDNA. The consensus sequences of the 135 declared satDNAs are
863 deposited in the NCBI GenBank under the accession numbers PQ553299-PQ553433. The satDNA
864 consensus sequences were BLAST-searched against NCBI GenBank database (Clark et al. 2016) using
865 megablast, discontiguous megablast, and blastn algorithms. The satDNA consensuses were screened
866 against GIRI Repbase, the database of repetitive DNA elements (Bao et al. 2015) using the CENSOR tool.
867

868 **SatDNA bioinformatic analyses**

869 The Geneious Prime 2023.2.1 package (Biomatters Ltd, New Zealand) was used for the basic analyses of
870 satellite repeats such as monomer length, A+T composition, mutual pairwise similarities, and potential
871 secondary structures. All other analyses were performed as described below, and the R scripts used in the
872 analyses are available as Supplementary Code at Figshare (figshare.com/s/d9b5c22dcbe842f27a8d).

873 *Repeat annotations*

874 The annotation of the *T. freemani* satDNA consensus sequences in the *T. freemani* genome assembly
875 Tfree1.0 (NCBI GenBank accession number GCA_939628115.1) and the *T. castaneum* genome assembly
876 TcasONT (ENA accession number GCA_950066185.1) was performed using NCBI's stand-alone BLAST
877 algorithm and the R programming language (R Core Team) with the metablastr package (Benoit and Drost
878 2021). The criteria for filtering the detected satDNAs in the genome were set to 70% percent identity and
879 70% query sequence coverage. The detected satDNA loci were converted to GFF files to facilitate
880 visualization and extraction. The annotations of satDNA repeats are available as Supplementary Data S6.

881 *Identification of satDNA arrays*

882 SatDNA arrays were detected by sorting all hits for a given satDNA in the BLAST results table and checking
883 whether another monomer of the same satDNA was found within a genomic distance of one monomer
884 length. If a monomer was detected, the array was extended and the step was repeated until the array
885 could no longer be extended (i.e. there were no repeats within a monomer length). This was particularly
886 applied when searching for satDNAs arrays with ≥5 consecutive repeats or scattered organization. The
887 annotations of satDNA arrays are available as Supplementary Data S6.

888 *Genomic environment*

889 Following array creation, the flanking regions of each array were examined for the presence of genes,
890 transposable elements and satDNAs discovered in the previous step. The used annotations for genes and

891 transposable elements were adopted from the official *T. freemani* annotation set available in FigShare at
892 10.6084/m9.figshare.19682400. The annotations for satDNAs are available as Supplementary Data S6. The
893 search space for hits was 10 kb +/- from the respective array start/end, and the statistics of resulting hits
894 were calculated.

895 *Organization of TfSat01 arrays*

896 The first step in analyzing the major satDNA TfSat01 was to revise its consensus sequence. Our TfSat01
897 consensus obtained by TAREAN analysis of WGS reads showed 9 nucleotide differences compared to
898 GenBank entry X58539 (**Suppl. Fig. S5A**), which was based on five randomly cloned monomers (Juan et al.
899 1993). We evaluated the two consensuses and concluded that our TfSat01 consensus is more
900 representative based on the following arguments. First, we BLAST-searched the two consensuses against
901 the Tfree1.0 assembly (using >70% similarity criterium) and detected 4.6% more monomer copies with the
902 TfSat01 consensus (228 060 for TfSat01, 218 083 for X58539). Second, the distribution of similarities
903 between the detected monomers and a query consensus showed a shift to the higher values for the
904 TfSat01 consensus (**Suppl. Fig. S5B**), suggesting that the TfSat01 sequence is more typical. Third, although
905 TfSat01 is derived as a consensus, it is not an artificial sequence. In the Tfree1.0 genome assembly, we
906 discovered 32 monomer copies that were 100% identical to the TfSat01 consensus, while no monomers
907 had 100% identity to X58539. We deposited the TfSat01 sequence as an updated consensus in the NCBI
908 GenBank database under the accession number PQ553299.

909 To visualize the centromeric organization of TfSat01, we used StainedGlass software (Vollger et al. 2022)
910 with the default settings and manually extracted centromeric regions of *T. freemani* chromosomes
911 comprising multi-megabase long TfSat01 arrays. To detect the inverted TfSat01 subarrays, the array
912 detection step was repeated in a strand-dependent manner to create the "strand-specific" database.
913 Subsequently, the "strand-specific" arrays were overlapped with total TfSat01 arrays (strand-independent)
914 to determine the number of inversions per strand-independent array. To determine the difference
915 between the lengths of "strand-specific" subarrays, the total length of the minus-strand subarrays was
916 deducted from the total length of the plus-strand subarrays and the differences were plotted. After
917 detection of TfSat01 arrays, the gaps among TfSat01 arrays were identified and examined for the presence
918 of the TfSat03 satDNA using overlaps and existing TfSat03 annotations. Following the detection of TfSat03
919 monomers in the gaps of TfSat01 arrays, the central regions where the TfSat03 repeats change orientation
920 were extracted and analyzed using graph networks as described below.

921 *Sequence alignments and visualizations*

922 All detected monomers of the 135 *T. freemani* satDNA were matched with the corresponding arrays to
923 assign them unique IDs and then extracted. To find orthologs in the *T. castaneum* genome, the same
924 processes of detection, array creation and extraction were repeated for *T. freemani* satDNAs in the
925 TcasONT assembly. Large multiple sequence alignments were generated using MAFFT (Katoh and Standley
926 2013). The genomic distance matrices were calculated using the dist.dna function with the "F81" genomic
927 distance model from the *ape* package (Paradis and Schliep 2019). Principal Component Analysis (PCA) was
928 subsequently performed on the distance matrix using the FactoMineR package (Lê et al. 2008) and
929 visualized with ggplot2 (static plots) and plotly (dynamic HTML plots). Network visualizations were created
930 from the distance matrices using the networkD3 package (available at https://cran.r-
931 project.org/web/packages/networkD3/ index.html), by finding the 5 closest neighbors of each sequence
932 in the alignment.

933

934 **Revision of fLG7 chromosome sequence in the *T. freemani* genome assembly Tfree1.0**

935 The reference genome of *T. freemani*, Tfree1.0 (GCA_939628115.1) comprises the sequences of nine
936 autosomes (fLG2-fLG10) and sex chromosome X (fLGX), but the sex chromosome y_p is not assembled
937 (Volarić et al. 2022). According to the Tfree1.0 assembly, the satDNA TfSat04 is located at the end of
938 chromosome fLG7, but in this work we established by *in situ* hybridization that the satDNA TfSat04 is y_p-
939 specific (**Fig. 4C**). Given the inconsistency between *in situ* and *in silico* localization of TfSat04, we revisited
940 the contigs that we used to build the Tfree1.0 assembly (Volarić et al. 2022). We found that the longest
941 continuous TfSat04 array is part of the ptg000052l contig, located distally on chromosome fLG7 and
942 connected to the rest of the chromosome by an $(N)_{100}$ placeholder representing an assembly gap (**Suppl.
943 Fig. S20**). The ptg000052l contig ends in TfSat04 repeats and has one of the lowest location confidence
944 scores in the assembly (Supplementary Table S3 in (Volarić et al. 2022)), so we hold that a slight similarity
945 to TfSat03 caused the contig to be misaligned to fLG7. After reviewing its sequence in detail, we conclude
946 that the ptg000052l contig is a candidate for the non-assembled y_p chromosome and support this
947 conclusion with several findings. First, in addition to the longest arrays of TfSat04, this contig also harbors
948 the longest array of the low-copy-number satDNA TfSat07 (**Suppl. Fig. S20**), and by *in situ* hybridization
949 we indeed detected the strongest TfSat07 FISH signal on the y_p chromosome (**Suppl. Fig. S3C**). Second,
950 given that assemblies in repetitive regions can be inaccurate, we examined the 24 Gb of the raw PacBio
951 HiFi reads used for the Tfree1.0 assembly to analyze gene dosage. Gene dosages for 32 genes annotated
952 on the ptg000052l contig and 2000 randomly selected genes annotated on different autosomes were
953 estimated by mapping the genes at the raw reads with minimap2 (Li 2018). Notably, 32 genes annotated
954 on the ptg000052l contig were represented in the raw reads on average two times less frequently

955 compared to 2000 randomly selected genes (**Suppl. Table S10**). Third, regarding the contig ends, one side
956 of the ptg000052l terminates in ~18 kb TfSat04 array, while the other side ends in a 1.2 kb $(TCAGG)_{243}$
957 array (**Suppl. Fig. S20**), which is the *Tribolium* telomeric sequence (Mravinac et al. 2011). The repetitive
958 endings suggest that the repeats at both ends extend further, but the ptg000052l contig with its size of
959 2.2 Mb probably accounts for the majority of the y_p sequence, the smallest chromosome of *T. freemani*.

960

961 **SatDNA probes**

962 To localize satDNAs on the chromosomes of *T. freemani* by fluorescence *in situ* hybridization (FISH) or to
963 prove their presence in the extrachromosomal circular DNA (eccDNA) fraction, specific DNA probes were
964 prepared. Specific probes were generated by PCR amplification of fragments of a satDNA of interest from
965 the genomic DNA, and subsequent cloning into a plasmid vector. A modified version of Primer3 2.3.7,
966 implemented in Geneious Prime 2023.2.1 package (Biomatters Ltd, New Zeland), was used to design
967 specific primers for each satDNA studied. The sequences of primer pairs and their optimal annealing
968 temperatures are specified in **Suppl. Table S11**. Reaction mixtures contained 10 ng genomic DNA, 0.2 µM
969 of each specific primer, 0.2 mM dNTP mix, 2.5 mM MgCl2, 0.25 U GoTaq G2 Flexi DNA polymerase, and 1x
970 Colorless GoTaq Flexi Buffer (Promega, USA). PCR amplification consisted of predenaturation at 94 ^{o}C for
971 3 min, 35 amplification cycles (denaturation at 94 ^{o}C for 10 s, annealing at optimal T_a for 10 s, extension at
972 72 ^{o}C for 10 s), and a final extension at 72 ^{o}C for 5 min. Cloning of satDNA fragments into the pGEM-T Easy
973 plasmid vector (Promega, USA) and transformation of *Escherichia coli* XL10-Gold Ultracompetent Cells
974 (Agilent Technologies, USA) were performed according to the manufacturer's instructions. Putative
975 positive clones were selected by blue-white color screening for recombinant plasmids, and insert lengths
976 were checked via colony-PCR amplification using the vector-specific primers M13F and M13R-40. Finally,
977 insert sequences of the clones used for satDNA probes were verified by Sanger sequencing (Macrogen
978 Europe BV, the Netherlands). To cover the variability of monomers within a satellite DNA, a mixture of
979 several clones was generally used to label each of the satDNA probes. SatDNA probes were labeled with
980 biotin-16-(5-aminoallyl)-dUTP or aminoallyl-dUTP-Cy3 (Jena Bioscience, Germany) by PCR amplification
981 using specific primers and cloned inserts as a template under conditions specified above. The ratio of
982 labeled dUTP and dTTP in the labeling reaction was 1:2.

983

984 **Chromosome preparations and fluorescence *in situ* hybridization (FISH)**

985 For FISH analyses, gonads isolated from *T. freemani* pupae were used to prepare chromosome spreads
986 using the squash method. Freshly isolated testes or ovaries were incubated in 10 µg/ml colcemid (Roche,
987 Switzerland) for 1 h. Hypotonic treatment was carried out in 0.075 M KCl for 5-15 min, followed by fixation

988 in a solution of absolute ethanol:acetic acid (3:1) for 15-30 min. The gonads were dissected on a slide in a
989 drop of 45% acetic acid, covered with a coverslip and squashed. Slides with coverslips were immersed in
990 liquid nitrogen for 30 s and the coverslips immediately removed with a razor blade. Slides were air-dried
991 and stored at -80 ^{o}C until use. Prior to *in situ* hybridization, slides with chromosome spreads were
992 subjected to the following treatments: incubation with RNase A (100 µg/mL) for 1 h at 37 ^{o}C, incubation
993 in 0.01% pepsin in 10 mM HCl for 10 min at 37 ^{o}C, incubation in 2.7% formaldehyde solution in PBS for 10
994 min at room temperature, dehydration through an ice-cold ethanol series (70% → 90% → 100%, 3 min
995 each), denaturation in 70% formamide at 70 ^{o}C for 2 min, and final dehydration through an ice-cold ethanol
996 series. Labeled satDNA probes (100-200 ng per slide) were first lyophilized (in double-color FISH
997 experiments, probes for different satDNAs were mixed before lyophilization), then dissolved in 15 µl of
998 hybridization buffer (60% formamide, 40% $DeSO_4$ buffer) and denatured at 75 ^{o}C for 5 min. After cooling
999 on ice, denatured probes were applied to chromosome slides and incubated in a moist chamber at 37 ^{o}C
1000 for overnight hybridization. Posthybridization washes were performed at 37 ^{o}C in 50% formamide in 2xSSC
1001 at 37 ^{o}C. While Cy3-labeled probes were ready for direct detection after posthybridization washes, biotin-
1002 labeled probes were visualized with fluorescein avidin D and biotinylated anti-avidin D system (Vector
1003 Laboratories, USA) using signal amplification. Signal amplification was achieved through three successive
1004 incubations in: 1) 1:500 dilution of fluorescein avidin D, 2) 1:100 dilution of biotinylated anti-avidin D, and
1005 3) 1:2000 dilution of fluorescein avidin D. Finally, slides were counterstained in 4',6-diamidino-2-
1006 phenylindole (DAPI) solution and embedded in Mowiol 4–88 mounting medium (Sigma-Aldrich, USA).
1007
1008 **Confocal microscopy and image analyses**
1009 Slides were examined with a Leica TCS SP8 X confocal laser scanning microscope, equipped with a HC PL
1010 APO CS2 63/1.40 oil objective, a 405 nm diode laser and a supercontinuum excitation laser (Leica
1011 Microsystems, Germany). Images were captured separately for each fluorochrome and processed with
1012 LASX Office 1.4.7 28921 (Leica Microsystems, Germany), ImageJ (Schneider et al. 2012) and Adobe
1013 Photoshop CS5 (Adobe Systems, USA), using only functions affecting the entire image equally. For each
1014 satDNA, at least 10 metaphase spreads from 3-10 independent experiments were analyzed.
1015
1016 **Extrachromosomal circular DNA isolation and two-dimensional agarose gel electrophoresis**
1017 To test the presence of satDNAs in the extrachromosomal circular DNA (eccDNA) molecules, total genomic
1018 DNA was isolated from 920 mg of mixed male and female larvae and adults of *T. freemani* according to the
1019 procedure described in (Volarić et al. 2021). DNA quantity was determined with a Qubit 4 fluorometer
1020 using Qubit dsDNA BR Assay Kit (Invitrogen, USA). In order to remove as much linear double-stranded DNA

1021 (dsDNA) as possible from the isolated DNA, the isolated DNA was treated mechanically and enzymatically.
1022 First, 20 µg of the isolated DNA was passed through a 0.33 mm syringe needle 30 times to shear the linear
1023 DNA. The sheared DNA was then treated with Exonuclease V, which degrades linear dsDNA in both 3' to
1024 5' and 5' to 3' directions leaving intact circular DNA. 20 µg of sheared DNA was digested with 200 U
1025 Exonuclease V (New England Biolabs, USA) at 37 ^{o}C overnight, and digestion was stopped by 11 mM EDTA
1026 pH 8.0 and incubation at 70 ^{o}C for 30 min. Prior to two-dimensional (2D) electrophoretic separation, DNA
1027 was cleaned using the Monarch® PCR & DNA Cleanup Kit (New England Biolabs, USA). Electrophoresis in
1028 the first dimension was performed in 0.7% ethidium bromide (EtBr)-free agarose gel in 1xTBE buffer at low
1029 voltage (0.7 V/cm) for 18 h. The first-dimension gel was then stained in 1x TBE with 0.3 µg/ml EtBr for 1 h.
1030 The entire gel strip of interest from the 1D agarose gel was excised, placed at a 90° angle to the direction
1031 of electrophoresis on a casting tray and poured with 1.5% agarose containing 0.3 µg/ml EtBr. The second-
1032 dimension electrophoresis was performed in 1xTBE containing 0.3 µg/ml EtBr at a higher voltage (4 V/cm)
1033 for 3 h. Both electrophoreses were run at room temperature.
1034
1035 **Southern blot hybridization**
1036 After 2D electrophoresis, agarose gels with separated DNA were agitated in 0.25 M HCl for 30 min and
1037 then in 0.4 M NaOH for 30 min. Depurinated and denatured DNA fragments were blotted onto a positively
1038 charged nylon membrane (Roche, Switzerland) overnight by capillary alkaline transfer in 0.4 M NaOH.
1039 Hybridization probes for TfSat01 and TfSat03 were prepared by PCR labeling using biotin-16-(5-
1040 aminoallyl)-dUTP (Jena Bioscience, Germany) and clones carrying specific satDNA fragments as described
1041 previously. Hybridization was performed in hybridization buffer (250 mM Na_2HPO_4 pH 7.2, 1 mM EDTA,
1042 20% SDS, 0.5% blocking reagent) with agitation overnight at 68 ^{o}C, while posthybridization washing was
1043 done in washing buffer (20 mM Na_2HPO_4, 1 mM EDTA, 1% SDS) at 65 ^{o}C. Chemiluminescent detection was
1044 carried out using the streptavidin-alkaline phosphatase (streptavidine-AP) conjugate and CDP-Star
1045 substrate (Roche, Switzerland). The signals were visualized using the Alliance Q9 Mini imaging system
1046 (Uvitec, UK). To hybridize the same blot with another probe, the first probe was stripped by washing the
1047 blot in stripping buffer (0.2 M NaOH, 0.1% SDS) at 42 ^{o}C for 2 x 15 min and then rinsed 3 x 5 min at room
1048 temperature in 2xSSC. After stripping and between hybridization experiments, blots were stored in 2xSSC
1049 at 4 ^{o}C to keep them wet. The stripping efficiency was tested by re-application of the streptavidin-AP
1050 conjugate and detection with the CDP-Star. To prove that stripping of the first probe does not affect
1051 hybridization with the second probe, multiple experiments were performed in which the order of the
1052 hybridization probes was changed.
1053

DATA AVAILABILITY

The raw Illumina sequencing data of the *T. freemani* genome are deposited in the NCBI Sequence Read Archive (SRA) database under the BioProject study accession number PRJNA1179347. The consensus sequences of the 135 *T. freemani* satDNAs are deposited in the NCBI GenBank under the accession numbers PQ553299-PQ553433. All scripts used for the analyses in this study are available as Supplementary Code at Figshare (figshare.com/s/d9b5c22dcbe842f27a8d). All supplementary figures, tables and data are available as Supplementary Material at Figshare (figshare.com/s/d9b5c22dcbe842f27a8d).

AUTHORS' CONTRIBUTIONS

Conceptualization - BM; Investigation - DV, EDS, MV, LH; Formal analysis - DV, EDS, MV, TVZ, BM; Visualization - DV, EDS, MV, BM; Resources - NM, BM; Funding acquisition and project administration - BM; Supervision - BM; Writing - original draft: BM; Writing review and editing - DV, EDS, MV, NM, BM. All authors read and approved the final manuscript.

FUNDING

This work has been fully supported by Croatian Science Foundation grant IP-2019-04-5522.

REFERENCES

Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* 6:4–9.

Belyayev A, Jandová M, Josefiová J, Kalendar R, Mahelka V, Mandák B, Krak K. 2020. The major satellite DNA families of the diploid Chenopodium album aggregate species: Arguments for and against the "library hypothesis." *PLoS One* 15:1–14.

Benoit M, Drost HG. 2021. A Predictive Approach to Infer the Activity and Natural Variation of Retrotransposon Families in Plants. *Methods Mol. Biol.* 2250:1–14.

Bersani F, Lee E, Kharchenko P V., Xu AW, Liu M, Xega K, MacKenzie OC, Brannigan BW, Wittner BS, Jung H, et al. 2015. Pericentromeric satellite repeat expansions through RNA-derived DNA intermediates in cancer. *Proc. Natl. Acad. Sci. U. S. A.* 112:15148–15153.

Boštjančić LL, Bonassin L, Anušić L, Lovrenčić L, Besendorfer V, Maguire I, Grandjean F, Austin CM, Greve C, Hamadou A Ben, et al. 2021. The Pontastacus leptodactylus (Astacidae) Repeatome Provides Insight Into Genome Evolution and Reveals Remarkable Diversity of Satellite DNA. *Front. Genet.* 11:611745.

1087 Brajković J, Feliciello I, Bruvo-Mađarić B, Ugarković D. 2012. Satellite DNA-like elements associated with genes
1088 within euchromatin of the beetle Tribolium castaneum. *G3 (Bethesda).* 2:931–941.

1089 Cabral-de-Mello DC, Mora P, Rico-Porras JM, Ferretti ABSM, Palomeque T, Lorite P. 2023. The spread of satellite
1090 DNAs in euchromatin and insights into the multiple sex chromosome evolution in Hemiptera revealed by
1091 repeatome analysis of the bug Oxycarenus hyalinipennis. *Insect Mol. Biol.* 32:725–737.

1092 Camacho JPM, Cabrero J, López-León MD, Martín-Peciña M, Perfectti F, Garrido-Ramos MA, Ruiz-Ruano FJ. 2022.
1093 Satellitome comparison of two oedipodine grasshoppers highlights the contingent nature of satellite DNA
1094 evolution. *BMC Biol.* 20:36.

1095 Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2016. GenBank. *Nucleic Acids Res.* 44:D67–D72.

1096 Cohen S, Yacobi K, Segal D. 2003. Extrachromosomal circular DNA of tandemly repeated genomic sequences in
1097 Drosophila. *Genome Res.* 13:1133–1145.

1098 Cohen Z, Bacharach E, Lavi S. 2006. Mouse major satellite DNA is prone to eccDNA formation via DNA Ligase IV-
1099 dependent pathway. *Oncogene* 25:4515–4524.

1100 Dias GB, Svartman M, Delprat A, Ruiz A, Kuhn GCS. 2014. Tetris is a foldback transposon that provided the building
1101 blocks for an emerging satellite DNA of Drosophila virilis. *Genome Biol. Evol.* 6:1302–1313.

1102 Elder JFJ, Turner BJ. 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. *Q. Rev. Biol.* 70:297–
1103 320.

1104 Feliciello I, Akrap I, Brajković J, Zlatar I, Ugarkovic D. 2015. Satellite DNA as a driver of population divergence in the
1105 red flour beetle tribolium castaneum. *Genome Biol. Evol.* 7:228–239.

1106 Feliciello I, Akrap I, Ugarković Đ. 2015. Satellite DNA Modulates Gene Expression in the Beetle Tribolium castaneum
1107 after Heat Stress. *PLoS Genet.* 11:e1005466.

1108 Ferree PM, Barbash DA. 2009. Species-specific heterochromatin prevents mitotic chromosome segregation to
1109 cause hybrid lethality in Drosophila. *PLoS Biol.* 7:e1000234.

1110 Flynn JM, Ahmed-Braimah YH, Long M, Wing RA, Clark AG. 2024. High-Quality Genome Assemblies Reveal
1111 Evolutionary Dynamics of Repetitive DNA and Structural Rearrangements in the Drosophila virilis Subgroup.
1112 *Genome Biol. Evol.* 16:1–15.

1113 Flynn JM, Yamashita YM. 2024. The implications of satellite DNA instability on cellular function and evolution.
1114 *Semin. Cell Dev. Biol.* 156:152–159.

1115 Fry K, Salser W. 1977. Nucleotide sequences of HS-α satellite DNA from kangaroo rat dipodomys ordii and
1116 characterization of similar sequences in other rodents. *Cell* 12:1069–1084.

1117 Gržan T, Despot-Slade E, Meštrović N, Plohl M, Mravinac B. 2020. CenH3 distribution reveals extended centromeres
1118 in the model beetle Tribolium castaneum. *PLoS Genet.* 16:e1009115.

1119 Gržan T, Dombi M, Despot-Slade E, Veseljak D, Volarić M, Meštrović N, Plohl M, Mravinac B. 2023. The Low-Copy-
1120 Number Satellite DNAs of the Model Beetle Tribolium castaneum. *Genes (Basel).* 14(5):999.

1121 Hartley G, O'Neill RJ. 2019. Centromere repeats: Hidden gems of the genome. *Genes (Basel).* 10(3):223.

1122 Henikoff S, Ahmad K, Malik HS. 2001. The centromere paradox: stable inheritance with rapidly evolving DNA.
1123 *Science* 293:1098–1102.

1124 Herndon N, Shelton J, Gerischer L, Ioannidis P, Ninova M, Dönitz J, Waterhouse RM, Liang C, Damm C, Siemanowski
1125 J, et al. 2020. Enhanced genome assembly and a new official gene set for Tribolium castaneum. *BMC*
1126 *Genomics* 21:47.

1127 Hinton HE. 1948. A synopsis of the genus Tribolium Macleay, with some remarks on the evolution of its species-
1128 groups (Coleoptera, Tenebrionidae). *Bull. Entomol. Res.* 39:13–55.

1129 Huang Y, Ding W, Zhang M, Han J, Jing Y, Yao W, Hasterok R, Wang Z, Wang K. 2021. The formation and evolution of
1130 centromeric satellite repeats in Saccharum species. *Plant J.* 106:616–629.

1131 Iwata T, Kishikawa T, Seimiya T, Notoya G, Suzuki T, Shibata CRPlosb pdfak., Miyakawa Y, Odawara N, Funato K,
1132 Tanaka E, et al. 2024. Satellite double-stranded RNA induces mesenchymal transition in pancreatic cancer by
1133 regulating alternative splicing. *J. Biol. Chem.* 300:105742.

1134 Jagannathan M, Yamashita YM. 2021. Defective Satellite DNA Clustering into Chromocenters Underlies Hybrid
1135 Incompatibility in Drosophila. *Mol. Biol. Evol.* 38:4977–4986.

1136 João Da Silva M, Gazoni T, Haddad CFB, Parise-Maltempi PP. 2023. Analysis in Proceratophrys boiei genome
1137 illuminates the satellite DNA content in a frog from the Brazilian Atlantic forest. *Front. Genet.* 14:1–11.

1138 Joseph A, Mitchell AR, Miller OJ. 1989. The organization of the mouse satellite DNA at centromeres. *Exp. Cell Res.*
1139 183:494–500.

1140 Juan C, Vazquez P, Rubio JM, Petitpierre E, Hewitt GM. 1993. Presence of highly repetitive DNA sequences in
1141 Tribolium flour-beetles. *Heredity (Edinb).* 70:1–8.

1142 Kalitsis P, Griffiths B, Choo KHA. 2006. Mouse telocentric sequences reveal a high rate of homogenization and
1143 possible role in Robertsonian translocation. *Proc. Natl. Acad. Sci. U. S. A.* 103:8786–8791.

1144 Kasinathan S, Henikoff S. 2018. Non-B-form DNA is enriched at centromeres. *Mol. Biol. Evol.* 35:949–962.

1145 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in
1146 performance and usability. *Mol. Biol. Evol.* 30:772–780.

1147 Kurdzo EL, Dawson DS. 2015. Centromere pairing - Tethering partner chromosomes in meiosis i. *FEBS J.* 282:2445–
1148 2457.

1149 Lê S, Josse J, Husson F. 2008. FactoMineR: An R Package for Multivariate Analysis. *J. Stat. Softw.* 25:1–18.

1150 Li H. 2018. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100.

1151 Lisachov A, Panthum T, Dedukh D, Singchat W, Ahmad SF, Wattanadilokcahtkun P, Thong T, Srikampa P, Noito K,
1152 Rasoarahona R, et al. 2024. Genome-wide sequence divergence of satellite DNA could underlie meiotic failure
1153 in male hybrids of bighead catfish and North African catfish (Clarias, Clariidae). *Genomics* 116:110868.

1154 Liu K, Wessler SR. 2017. Transposition of Mutator-like transposable elements (MULEs) resembles hAT and Transib
1155 elements and V(D)J recombination. *Nucleic Acids Res.* 45:6644–6655.

1156 Liu Q, Yi C, Zhang Z, Su H, Liu Chang, Huang Y, Li W, Hu X, Liu Cheng, Birchler JA, et al. 2023. Non-B-form DNA tends
1157 to form in centromeric regions and has undergone changes in polyploid oat subgenomes. *Proc. Natl. Acad.*
1158 *Sci. U. S. A.* 120:e2211683120.

1159 Mellone BG, Fachinetti D. 2021. Diverse mechanisms of centromere specification. *Curr. Biol.* 31:R1491–R1504.

1160 Meštrović N, Mravinac B, Pavlek M, Vojvoda-Zeljko T, Šatović E, Plohl M. 2015. Structural and functional liaisons
1161 between transposable elements and satellite DNAs. *Chromosom. Res.* 23:583–596.

1162 Meštrović N, Plohl M, Mravinac B, Ugarković Đ. 1998. Evolution of satellite DNAs from the genus Palorus -
1163 experimental evidence for the “library” hypothesis. *Mol. Biol. Evol.* 15:1062–1068.

1164 Mravinac B, Meštrović N, Vanja Čavrak V, Plohl M. 2011. TCAGG, an alternative telomeric sequence in insects.
1165 *Chromosoma* 120:367–376.

1166 Mravinac B, Plohl M. 2010. Parallelism in evolution of highly repetitive DNAs in sibling species. *Mol. Biol. Evol.*
1167 27:1857–1867.

1168 Mravinac B, Ugarković D, Franjević D, Plohl M. 2005. Long inversely oriented subunits form a complex monomer of
1169 Tribolium brevicornis satellite DNA. *J. Mol. Evol.* 60:513–525.

1170 Nakakita H, Imura O, Winks RG. 1981. Hybridization between Tribolium freemani Hinton and Tribolium castaneum
1171 (Herbst), and Some Preliminary Studies on the Biology of Tribolium freemani (Coleoptera: Tenebrionidae).
1172 *Appl. Entomol. Zool.* 16:209–215.

1173 Navrátilová A, Koblížková A, Macas J. 2008. Survey of extrachromosomal circular DNA derived from plant satellite
1174 repeats. *BMC Plant Biol.* 8:1–13.

1175 Novák P, Robledillo LÁ, Koblížková A, Vrbová I, Neumann P, Macas J. 2017. TAREAN: A computational tool for
1176 identification and characterization of satellite DNA from unassembled short reads. *Nucleic Acids Res.* 45:e111.

1177 Packiaraj J, Thakur J. 2024. DNA satellite and chromatin organization at mouse centromeres and pericentromeres.
1178 *Genome Biol.* 25:1–24.

1179 Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R.
1180 *Bioinformatics* 35:526–528.

1181 Pavlek M, Gelfand Y, Plohl M, Meštrović N. 2015. Genome-wide analysis of tandem repeats in Tribolium castaneum
1182 genome reveals abundant and highly dynamic tandem repeat families with satellite DNA features in
1183 euchromatic chromosomal arms. *DNA Res.* 22:387–401.

1184 Pertile MD, Graham AN, Choo KHA, Kalitsis P. 2009. Rapid evolution of mouse Y centromere repeat DNA belies
1185 recent sequence stability. *Genome Res.* 19:2202–2213.

1186 Ramesh B, Firneno TJ, Demuth JP. 2021. Divergence time estimation of genus Tribolium by extensive sampling of
1187 highly conserved orthologs. *Mol. Phylogenet. Evol.* 159:107084.

1188 Richards S, Gibbs RA, Weinstock GM, Brown S, Denell R, Beeman RW, Gibbs R, Bucher G, Friedrich M,
1189 Grimmelikhuijzen CJP, et al. 2008. The genome of the model beetle and pest Tribolium castaneum. *Nature*
1190 452:949–955.

1191 Rico-Porras JM, Mora P, Palomeque T, Montiel EE, Cabral-de-Mello DC, Lorite P. 2024. Heterochromatin Is Not the
1192 Only Place for satDNAs: The High Diversity of satDNAs in the Euchromatin of the Beetle Chrysolina americana
1193 (Coleoptera, Chrysomelidae). *Genes (Basel).* 15 (4):395.

1194 Schmidt N, Sielemann K, Breitenbach S, Fuchs J, Pucker B, Weisshaar B, Holtgräwe D, Heitkam T. 2024. Repeat
1195 turnover meets stable chromosomes: repetitive DNA sequences mark speciation and gene pool boundaries in
1196 sugar beet and wild beets. *Plant J.* 118:171–190.

1197 Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods*
1198 9:671–675.

1199 Shimeld L. 1989. A cytogenetic examination of eight species of Tribolium. *Tribolium Inf. Bull.* 29:102–107.

1200 Silva DMZ de A, Utsunomia R, Ruiz-Ruano FJ, Daniel SN, Porto-Foresti F, Hashimoto DT, Oliveira C, Camacho JPM,
1201 Foresti F. 2017. High-throughput analysis unveils a highly shared satellite DNA library among three species of
1202 fish genus Astyanax. *Sci. Rep.* 7:12726.

1203 Sokoloff A. 1972. The Biology of Tribolium, With Special Emphasis on Genetic Aspects, Vol. 1. Clarendon Press,
1204 Oxford

1205 Spangenberg V, Losev M, Volkhin I, Smirnova S, Nikitin P, Kolomiets O. 2021. DNA environment of centromeres and
1206 non-homologous chromosomes interactions in mouse. *Cells* 10:1–17.

1207 Sproul JS, Khost DE, Eickbush DG, Negm S, Wei X, Wong I, Larracuente AM. 2020. Dynamic evolution of euchromatic

1208 satellites on the x chromosome in drosophila melanogaster and the simulans clade. *Mol. Biol. Evol.* 37:2241–
1209 2256.

1210 Talbert PB, Henikoff S. 2022. The genetics and epigenetics of satellite centromeres. *Genome Res.* 32:608–615.

1211 Ugarković D, Durajlija S, Plohl M. 1996. Evolution of Tribolium madens (Insecta, Coleoptera) satellite DNA through
1212 DNA inversion and insertion. *J. Mol. Evol.* 42:350–358.

1213 Ugarković D, Podnar M, Plohl M. 1996. Satellite DNA of the red flour beetle Tribolium castaneum - Comparative
1214 study of satellites from the genus Tribolium. *Mol. Biol. Evol.* 13:1059–1066.

1215 Volarić M, Despot-Slade E, Veseljak D, Meštrović N, Mravinac B. 2022. Reference-Guided De Novo Genome
1216 Assembly of the Flour Beetle Tribolium freemani. *Int. J. Mol. Sci.* 23(11):5869.

1217 Volarić M, Despot-Slade E, Veseljak D, Mravinac B, Meštrović N. 2024. Long-read genome assembly of the insect
1218 model organism Tribolium castaneum reveals spread of satellite DNA in gene-rich regions by recurrent burst
1219 events. *Genome Res.* 34:1878–1894.

1220 Volarić M, Veseljak D, Mravinac B, Meštrović N, Despot-Slade E. 2021. Isolation of High Molecular Weight DNA from
1221 the Model Beetle Tribolium for Nanopore Sequencing. *Genes (Basel).* 12(8):1114.

1222 Vollger MR, Kerpedjiev P, Phillippy AM, Eichler EE. 2022. StainedGlass: interactive visualization of massive tandem
1223 repeat structures with identity heatmaps. *Bioinformatics* 38:2049–2051.

1224 Vondrak T, Avila Robledillo L, Novak P, Koblizkova A, Neumann P, Macas J. 2020. Characterization of repeat arrays
1225 in ultra-long nanopore reads reveals frequent origin of satellite DNA from retrotransposon-derived tandem
1226 repeats. *Plant J.* 101:484–500.

1227 Wade MJ, Johnson NA. 1994. Reproductive isolation between two species of flour beetles, Tribolium castaneum
1228 and T. freemani: Variation within and among geographical populations of T. castaneum. *Heredity (Edinb).*
1229 72:155–162.

1230 Zattera ML, Bruschi DP. 2022. Transposable Elements as a Source of Novel Repetitive DNA in the Eukaryote
1231 Genome. *Cells* 11:3373.

1232 Žinić SD, Ugarković D, Cornudella L, Plohl M. 2000. A novel interspersed type of organization of satellite DNAs in
1233 Tribolium madens heterochromatin. *Chromosom. Res.* 8:201–212.

1234

1235

1236

1237

1238

1239 **SUPPLEMENTARY MATERIAL**

1240 **Supplementary Figure S1.** Consensus sequence alignments of satDNAs belonging to four superfamilies
1241 (SF1-SF4). The attached matrices show the percentage of pairwise identity between the compared
1242 sequences. For the superfamily SF1 (**A**), the comparison between the satellite TfSat04 and the three
1243 subunits belonging to the satDNA TfSat03 is presented.

1244 **Supplementary Figure S2.** Alignments of the subunits of satDNA TfSat03 (**A**), TfSat39 (**B**), TfSat41 (**C**),
1245 TfSat76 (**D**), and TfSat77 (**E**), whose monomers are based on higher-order-repeats (HORs). Schematic
1246 representations show the HOR structure, the lengths of the subunits and their pairwise similarities.

1247 **Supplementary Figure S3.** Localization of TfSat05 (**A**), TfSat06 (**B**), TfSat07 (**C**), TfSat08 (**D**), TfSat09 (**E**) and
1248 TfSat10 (**F**) satDNAs on *T. freemani* chromosomes (2n=20), determined by fluorescence *in situ*
1249 hybridization. Chromosomes were stained with 4',6-diamidino-2-phenylindole (DAPI) and hybridized with
1250 satDNA probes visualized by fluorescein isothiocyanate (FITC). For each satDNA localization, chromosomes
1251 are shown in black-and-white (first panel) and in DAPI-stained version (second panel). The position of a
1252 satDNA-specific probe is indicated by green fluorescence in the third panel, while the fourth panel shows
1253 a merge of DAPI-stained chromosomes and FITC-vizualized satDNA probes. For satDNA TfSat07 (**C**), the
1254 position of the signals on the male chromosome y_p is indicated by an arrow. The scale bar corresponds to
1255 3 µm.

1256 **Supplementary Figure S4.** Distances of genes, satellite DNAs (satDNAs) and transposable elements (TEs)
1257 to *T. freemani* satDNAs arrays (containing ≥5 consecutive monomers) in their 10-kb flanking regions. Each
1258 dot represents the mean distance of annotated sequences of interest for each satDNA. The x-axis indicates
1259 the mean position value in the 10-kb flanking region, with negative and positive values indicating the left
1260 and right flanking regions, respectively. The mean distance values were calculated based on the data
1261 provided for individual arrays in Supplementary Table S5.

1262 **Supplementary Figure S5.** Comparison of the consensus sequences TfSat01 (defined in this work) and
1263 GenBank entry X58539 (Juan *et al.*, 1993). **A)** The alignment of the consensus sequences. The nucleotide
1264 differences between the compared consensuses are colored, and dashes indicate alignment gaps. **B)**
1265 Distribution of nucleotide sequence similarities between monomers annotated in the *T. freemani* genome
1266 assembly Tfree1.0 and the consensus sequence TfSat01 (left panel) and the consensus X58539 (right
1267 panel). The BLAST search was done by using the criteria of >70% sequence similarity and >70% query
1268 coverage. The black arrow points to the monomer group that has 82% similarity to the TfSat01 consensus.

1269 **Supplementary Figure S6.** TfSat01 multi-megabase regions in the Tfree1.0 assembly visualized with
1270 StainedGlass (Vollger *et al.*, 2022). The StainedGlass analysis was performed using 10 kb windows, and the
1271 sequence identity values are shown in the color histogram along each sequence identity heatmap. Two
1272 discovered multi-megabase regions (region 1 and region 2) are shown for chromosome fLG9. The lengths
1273 and positions of the multi-megabase regions are listed in Supplementary Table S8.

1274 **Supplementary Figure S7.** Density plots of the differences between the length of the direct and inverted
1275 subarray within the 1793 TfSat01 arrays distributed at seven *T. freemani* chromosomes (fLG2-fLG9). The
1276 X-axis shows the difference in the length of directly and inversely oriented subarrays within an array. The
1277 relative abundance of subarray length differences in the graph is indicated by color gradient.

1278 **Supplementary Figure S8.** Organization of the six most frequent inversion segments (26 bp, 106 bp, 132
1279 bp, 149 bp, 187 bp, and 200 bp long) in which TfSat01 monomers change their orientation within the
1280 TfSat01 arrays. **A)** A schematic showing the position of the inversely oriented truncated TfSat01 monomers

in the inversion segments. The labels "DIR" and "INV" indicate the direct and inverted orientation of the truncated monomers within the inversion segment. **B)** Alignment of the truncated TfSat01 monomers in the inversion segments. The sequence "TfSat01_direct_vs_inverted" represents directly and inversely oriented TfSat01 consensus monomers, according to which the truncated monomers of the inversion site are aligned. **C)** Potential secondary structures of six inversion segments predicted by the RNAfold tool (Lorenz *et al*., 2011) using DNA Matthews 1999 energy model. The minimum free energies of the structures are given in parentheses. **D)** List of chromosomes on which a particular inversion segment is present. The alignments of sequences belonging to the particular inversion segments are shown in Supplementary Data S1.

Supplementary Figure S9. The sequence of the central region in the intercalary segment where the degenerate TfSat03 copies change orientation. **A)** Schematic representation of the central region with the inverted segments marked by red and green arrows. **B)** Potential secondary structure of the central region predicted by the RNAfold tool (Lorenz *et al*., 2011) using DNA Matthews 1999 energy model (the minimum free energy of the structures = - 6.06 kcal/mol).

Supplementary Figure S10. Graph network of ~110 bp long central regions from 1577 intercalary segments in which TfSat03 monomers change their orientation. The graph illustrates the sequence similarity relationships based on the alignment presented in Supplementary Data S2. Each dot represents one sequence, and the color of the dot indicates the chromosome on which the sequence is located. The color legend indicates individual chromosomes. The cluster, which consists of 100% identical sequences originating from different chromosomes, is marked with a black square and shown enlarged in Figure 3D. An interactive representation of the graph network is given in the Supplementary Data S3.

Supplementary Figure S11. The alignment between TfSat01, TfSat02 and TCsat15 consensus sequences. Identities between the compared sequences are indicated by dots (.), and dashes (-) indicate alignment gaps. The box indicates the 166 bp segment within the TfSat02 and TCsat15 consensuses that corresponds to the TfSat01 sequence. The blue arrow shows the position and orientation of the TfSat01 sequence.

Supplementary Figure S12. TCAST transposon-like element structure and genome distribution. **A)** Schematic representation of the TCAST transposon-like element with the indicated positions of a 367 bp long segment corresponding to the TCAST satDNA monomer (TCAST) and its 143 bp long truncated version (TCAST***). The inverted orientation of the segments is indicated by the direction of the red arrows. The terminal inverted repeats (TIR) of the DNA transposon-like element are indicated by gray arrows. **B)** Distribution of the 57 copies of the TCAST transposon-like element annotated in the *T. castaneum* assembly TcasONT (upper panel) and 84 copies annotated in the *T. freemani* genome assembly Tfree1.0 (lower panel). **C)** Maximum likelihood tree of the individual copies of the TCAST transposon-like element detected in the *T. castaneum* assembly (red) and in the *T. freemani* assembly (blue). The tree was inferred in iTOL with nodal supports based on 100 bootstrap replicates (the values ≥80 are shown).

Supplementary Figure S13. Relationships between *T. freemani* satDNAs TfSat03 and TfSat04, and *T. castaneum* satDNA Cast7. In the consensus sequences' alignment of Cast7 and TfSat04 monomers and TfSat03 subunits A, B and C, differences between the compared consensuses are colored and dashes indicate alignment gaps. The attached matrix shows the percentage of pairwise identity between the compared sequences.

Supplementary Figure S14. Alignments of the satDNAs *T. freemani* satDNAs (TfSat) and their orthologs among the known *T. castaneum* satDNAs (Cast and TCsat). Pairwise identities between the consensus sequences are indicated in the parentheses.

1324 **Supplementary Figure S15.** Relationships between the *T. freemani* satDNAs and their *T. castaneum*
1325 orthologs associated with DNA transposons from the Rehavkus superfamily. **A)** Phylogenetic relationships
1326 between the TfSat15 and TCsat23 repeats present in the inverted termini of the Rehavkus-1_TC DNA
1327 transposon. The schematic of the Rehavkus-1_TC is shown in the center, while the relationships between
1328 the *T. freeemani*/*T. castaneum* satDNA repeats from the inverted termini and the relationships between
1329 the central part of the transposon sequence are shown in the upper and lower panels, respectively. The
1330 sequences of *T. freemani* are represented by blue dots, while red dots indicate *T. castaneum* sequences.
1331 The maximum likelihood trees were inferred in iTOL with nodal supports based on 100 bootstrap
1332 replicates. **B)** Phylogenetic relationships between the TfSat23 and its orthologous *T. castaneum* repeats
1333 present in the inverted termini of the Rehavkus-3_TC DNA transposon. The same description of the tree
1334 applies as for the previous one. **C)** The principal component analysis (PCA) of the TfSat25 repeats present
1335 in the Rehavkus-like element in *T. freemani* and the orthologous repeats of the *T. castaneum* satDNA
1336 TCsat12 present in the 80.5 kb long array on *T. castaneum* chromosome LG4. In the PCA plot, the *T.*
1337 *freemani* repeats are indicated by blue dots, while the *T. castaneum* repeats are indicated by red dots.
1338 Schematic on the right illustrates the different organization of the orthologous repeats in the two sibling
1339 species.

1340 **Supplementary Figure S16.** Comparison of the 10-kb flanking regions of the orthologous satDNAs TfSat08
1341 and TCsat17 reveals the syntenic locus where the satDNAs are located in *T. freemani* (on chromosome
1342 fLG3) and *T. castaneum* (on chromosome LG3). The interspecific sequence similarities between the
1343 flanking regions are indicated within up-down arrows.

1344 **Supplementary Figure S17.** Organization of the *T. freemani* satDNA TfSat06. **A)** The principal component
1345 analysis of the TfSat06 repeats from different *T. freemani* chromosomes. Dots represent repeats colored
1346 according the chromosome (fLG) of origin. **B)** Schematic representation of the organization of the
1347 Rehavkus-like element whose inverted termini harbor the TfSat06 repeats (represented by green arrows).
1348 The table reports the segmental similarities of the central part of the Rehavkus-like element with the
1349 repetitive elements deposited in the Repbase database. The similarities refer to the segments (purple
1350 blocks in the diagram) whose positions are listed in the table together with the similarities and alignment
1351 scores.

1352 **Supplementary Figure S18.** Principal component analysis (PCA) of the orthologous satDNAs TfSat22 and
1353 Cast9. The upper panel shows intense mixing of *T. freemani* TfSat22 repeats (blue dots) and *T. castaneum*
1354 Cast9 repeats (red dots). Interchromosomal homogenization within the species is shown in the lower
1355 panels (*T. freemani* left, *T. castaneum* right), with the repeats of individual chromosomes colored
1356 according to the attached color legend.

1357 **Supplementary Figure S19.** Organization of the *T. freemani* low-copy-number satDNAs TfSat54 (**A**) and
1358 TfSat97 (**B**), showing intraarray homogenization. The positions of the arrays on the chromosomes are
1359 indicated on the schematics (the size of the chromosomes is indicated on the left). Clusters formed by
1360 repeats of the same array are circled in the PCA plots. Red dots represent orthologous copies detected in
1361 the sibling species *T. castaneum*.

1362 **Supplementary Figure S20.** The position of the ptg000052l contig (green rectangle in the center panel) on
1363 the *T. freemani* chromosome fLG7 in the Tfree1.0 assembly. The top panel shows an enlarged view of the
1364 junction between the pg000052L contig and the rest of the fLG7 chromosome sequence through the $(N)_{100}$
1365 placeholder. The lower panel shows the ptg000052l contig with the indicated positions of satDNA TfSat04
1366 (magenta arrows), satDNA TfSat07 (brown arrows) and telomeric repeats TCAGG (dark green arrows).

1367 **Supplementary Table S1.** The output of six clustering analyses (T1-T6) performed by the TAREAN pipeline
1368 using Illumina paired-end whole genome sequence reads of the flour beetle *T. freemani*. With the
1369 exception of analysis T1, all other analyzes (T2-T6) used sets of input reads from which reads
1370 corresponding to the major satDNA (Juan *et al*., 1993) were excluded. The genome coverage was
1371 calculated based on the estimated T. freemani genome size of 320 Mb (Volarić *et al*., 2022).

1372 **Supplementary Table S2.** Consensus sequences and GenBank accession numbers of the 135 *T. freemani*
1373 satDNAs identified in this work.

1374 **Supplementary Table S3.** Main characteristics of the 135 *T. freemani* satDNAs identified in this work. The
1375 last two columns specify orthologous satDNAs from the sibling species *T. castaneum* found by BLAST
1376 searches against the NCBI GenBank database and repetitive elements from the Repbase database with
1377 which *T. freemani* satDNAs showed partial similarities when searched with the CENSOR tool.

1378 **Supplementary Table S4.** Number of satDNA monomers annotated in the Tfree1.0 genome assembly using
1379 >70% sequence similarity criterion.

1380 **Supplementary Table S5.** Distance of nearest annotated genes, satellite DNAs (satDNA) and transposable
1381 elements (TE) in the flanking 10-kb regions (left and right) for individual arrays of the *T. freemani* satDNAs.
1382 The arrays are defined as stretches containing ≥5 consecutive monomers. The array ID name indicates the
1383 fLG chromosome (beginning of the name) and the position on the chromosome (end of the name) on
1384 which the array is located. Positions and distances are given according to the annotations in the Tfree1.0
1385 genome assembly (Volarić *et al*., 2022).

1386 **Supplementary Table S6.** Summary data on the presence of genes, satDNAs and transposable elements
1387 (TE) in the 10-kb flanking regions of the *T. freemani* satDNAs. The summary data are based on the data
1388 given for individual arrays in Supplementary Table S5. "Yes" indicates that at least one array has an
1389 annotated sequence of interest in its 10-kb surrounding region (left or right), whereas "No" indicates that
1390 no array has an annotated sequence of interest in the 10-kb surrounding regions.

1391 **Supplementary Table S7.** Average distance (mean and median) of genes, satellite DNAs (satDNAs) and
1392 transposable elements (TEs) from *T. freemani* satDNA arrays (with ≥5 consecutive monomers) analyzed in
1393 their 10-kb flanking regions (left and right). Average distances were calculated based on the data given for
1394 individual arrays in Supplementary Table S5.

1395 **Supplementary Table S8.** Lengths and positions of the TfSat01 multi-megabase regions on individual
1396 chromosomes in the Tfree1.0 assembly.

1397 **Supplementary Table S9.** The number of the subarrays within a continuous TfSat01 array. The second
1398 column indicates the number of TfSat01 arrays annotated in the Tfree1.0 assembly, which contain the
1399 number of subarrays specified in the first column.

1400 **Supplementary Table S10.** Quantitates of 32 genes from the ptg000052L contig and 2000 randomly
1401 selected genes from the Tfree1.0 assembly in 24 GB of raw data obtained by PacBio HiFi sequencing of the
1402 *T. freemani* genome. The occurrence of the genes was determined by mapping their sequences against
1403 HiFi reads using the minimap2 tool.

1404 **Supplementary Table S11.** The sequences of *T. freemani* specific satDNA primers used in this study. The
1405 optimal annealing temperature (T_a) for each primer pair is indicated.

1406 **Supplementary Data S1.** Alignments of segments in which TfSat01 monomers change their orientation
1407 within the TfSat01 arrays. Shown are the alignments for the six most common segments (26 bp, 106 bp,
1408 132 bp, 149 bp, 187 bp and 200 bp long). The names of the aligned sequences indicate the *T. freemani* fLG
1409 chromosome (beginning of the name) and the position on the chromosome (end of the name) on which
1410 the sequence is located.

1411 **Supplementary Data S2.** The alignment of 1577 sequences of the ~110 bp long central region of intercalary
1412 segments in which TfSat03 monomers change their orientation. The names of the aligned sequences
1413 indicate the *T. freemani* fLG chromosome (beginning of the name) and the position on the chromosome
1414 (end of the name) on which the sequence is located.

1415 **Supplementary Data S3.** Interactive graph network of ~110 bp long central regions from 1577 intercalary
1416 segments in which TfSat03 monomers change their orientation. The graph illustrates the sequence
1417 similarity relationships based on the alignment presented in Supplementary Data S2. Each dot represents
1418 one sequence, and the color of the dot indicates the *T. freemani* fLG chromosome on which the sequence
1419 is located. The color legend indicates individual chromosomes. The position of the sequence on the
1420 chromosome is revealed by touching a dot with the mouse pointer. The relationships between the
1421 individual sequences within the network can be examined by zooming in/out.

1422 **Supplementary Data S4.** Alignments of *T. freemani* satDNAs (TfSat) monomers annotated in the *T.*
1423 *freemani* Tfree1.0 genome assembly and their orthologous copies annotated in the *T. castaneum* TcasONT
1424 assembly. For each annotated monomeric repeat, the beginning of the name (Tfree/Tcas) indicates the
1425 species in whose assembly the repeat is annotated. The designations fLG and LG in the names of the
1426 repeats indicate the chromosome and the position at which the repeat is annotated (fLG for *T. freemani*
1427 chromosomes, LG for *T. castaneum* chromosomes).

1428 **Supplementary Data S5.** Principal component analyses (PCAs) of aligned *T. freemani* satDNAs (TfSat)
1429 repeats annotated in the *T. freemani* Tfree1.0 genome assembly and their orthologous copies annotated
1430 in the *T. castaneum* TcasONT assembly. The PCA plots are based on the alignments presented in
1431 Supplementary Data S4. The dots represent monomer repeats colored according to the species of origin
1432 (*T. freemani* in blue, *T. castaneum* in red). By moving the mouse pointer over a dot, the chromosomal
1433 location of the annotated monomer is displayed indicating a chromosome (fLG for *T. freemani*, LG for *T.*
1434 *castaneum*) and repeat's position on a chromosome.

1435 **Supplementary Data S6.** Annotations of *T. freemani* stDNA monomers and arrays annotated in the *T.*
1436 *freemani* genome assembly Tfree1.0.

1437