Utjecaj satelitne DNA na regulaciju gena u brašnara Tribolium castaneum (Herbst 1797)

Akrap, Ivana

University of Zagreb
Faculty of Science
Department of Biology

Ivana Akrap

Study of satellite DNA-mediated gene regulation in red flour beetle *Tribolium castaneum* (Herbst, 1797)

Graduation Thesis

Zagreb, 2013

This thesis was made at Ruđer Bošković Institute, under supervision of prof. Dr. Đurđica Ugarković and dr. Isidoro Feliciello from University of Naples Federico II, and submitted for evaluation to Department of Biology, Faculty of Science in order to acquire title the Master of molecular biology.

BASIC DOCUMENTATION CARD

University of Zagreb

Faculty of Science

Department of Biology Graduation Thesis

Study of of satellite DNA-mediated gene regulation in red flour beetle *Tribolium castaneum* (Herbst, 1797)

Ivana Akrap

Rooseveltov trg 6, 10000 Zagreb, Croatia

Satellite DNAs are tandemly repeated sequences that constitute a considerable part of the genomic DNA of many eukaryotic organisms. They represent a major building element of pericentromeric and centromeric heterochromatin. The existence of conserved motifs and structural properties and their transcriptional activity suggest that, in addition to participating in centromere and heterochromatin formation, satellite DNAs might also act as *cis*-regulatory elements of gene expression. TCAST represents major satellite DNA in the red flour beetle *Tribolium castaneum.* Dispersed form of TCAST satellite was associated with 101 protein-coding genes. This study reveals polymorphism of dispersed TCAST satellite-like elements among strains of *T. castaneum* and gives a strong indication for the role of site specific recombination as a mechanism of dispersion of repetitive elements. The results of this study also suggest the role of dispersed TCAST satellite-like elements in regulation of genes located in their vicinity, under specific heat stress conditions.

(52 pages, 17 figures, 3 tables, 44 references, original in: English)

Thesis deposited in the Central Biological Library

Key words*: cis*-regulatory elements, satellite polymorphism, TCAST satellite

Supervisor: Đurđica Ugarković, PhD, Adjunct Prof.

Reviewers: Mirjana Kalafatić, PhD, Prof., Vlatka Zoldoš, PhD, Assoc. Prof., Dijana Škorić, PhD, Assoc. Prof.

Thesis accepted: 24.1 2013.

TEMELJNA DOKUMENTACIJSKA KARTICA

Sveučilište u Zagrebu

Prirodoslovno-matematički fakultet

Biološki odsjek Diplomski rad

Utjecaj satelitne DNA na regulaciju gena u brašnara *Tribolium castaneum* (Herbst, 1797)

Ivana Akrap

Rooseveltov trg 6, 10000 Zagreb

Satelitne DNA predstavljaju uzastopno ponavljajuće sekvence koje čine značajan dio genoma eukariotskih organizama. Glavni su građevni elementi pericentromernog i centromernog hetrokromatina. Njihove strukturne odlike, očuvani motivi i transkripcijska aktivnost upućuju također na njihovu ulogu u regulaciji ekspresije gena. TCAST predstavlja glavnu satelitnu DNA u brašnara *Tribolium castaneum*. Raspršeni oblik TCAST-a je pronađen u blizini velikog broja (101) unikatnog gena. U ovome radu otkriven je polimorfizam raspršenog satelita TCAST među sojevima brašnara *T. castaneum* i kao mogući mehanizam raspršenja satelita predložena je mjesno-specifična rekombinacija. Dobiveni rezultati također upućuju na moguću ulogu raspršenog satelita TCAST u regulaciji gena koji se nalaze u njegovoj blizini, u uvjetima temperaturnog stresa.

(52 stranica, 17 slika, 3 tablice, 44 literaturnih navoda, jezik izvornika: engleski)

Rad je pohranjen u Središnjoj biološkoj knjižnici

Ključne riječi: *cis*- regulatorni elementi, polimorfizam satelitnih sekvenci, satelit TCAST

Voditelj: Dr. sc. Đurđica Ugarković , nasl. red. prof.

Ocjenitelji: Dr. sc. Mirjana Kalafatić, red. prof., Dr. sc. Vlatka Zoldoš, izv. prof., Dr. sc. Dijana Škorić, izv. prof.

Rad prihvaćen: 24.1. 2013.

Contents

1. Introduction

1.1.Tribolium castaneum

Red flour beetle *Tribolium castaneum* is tenebrionid beetle. Its size can variet from 2.3 to 4.4 mm. It has 4 developmental stages: eggs, larvae, pupa and adult. *Tribolium* generation time is temperature dependant and can take from 20 days on 37.5 °C to more then 140 days < 20 °C. They can prosper on wide variety of grain, cereal and nut products at >10% relative humidity (Howe, 1956). *Tribolium castaneum* is a major global pest in the agricultural industry, causing billions of dollars worth losses on stored grain and cereal products. *Tribolium castaneum* provides an excellent genetic model system for Coleopterans, the largest and most diverse order of eukaryotic organisms. (Brown et al., 2003). Similar to *Drosophila melanogaster* in the order Diptera, *Tribolium castaneum* has characteristics of a typical genetic model organism: ease of culture, short generation time, large brood sizes and efficacy of genetic manipulation. The potential of *Tribolium castaneum* for genetic analysis has been demonstrated through RNA interference (Tomoyasu et al., 2008), whole-genome molecular mapping and classical mutational studies (Lorenzen et al., 2005). Completion of the genome sequence in 2008 has greatly facilitated molecular genetics and genomic studies in *Tribolium castaneum*. Sequencing involved the euchromatic portion of the genome, with >20% of the genome, corresponding to heterochromatic regions, excluded due to technical difficulties. The sequence data provides useful information for identifying and characterizing the function and organization of beetle genes as well as their orthologues in other insect species. *Tribolium castaneum* is momentarily the most efficient model system for performing functional analysis of genes lost in the Drosophila lineage but conserved in other insects (Tribolium Genome Sequencing Consortium, 2008).

1.2. Satellite DNA

Satellite DNAs (satDNAs) are tandemly repeated sequences that constitute a considerable part of the genomic DNA of eukaryotic organism as they are major component of genetically silent heterochromatin. They are generally formed by long tandem arrays in which the monomers are repeated in head-to-tail fashion predominantly. satDNAs are located in pericentromeric, centromeric and/or telomeric regions of chromosome. (Palomeque and Lorite, 2008).

Despite the fact, that they are major building element of pericentromeric and centromeric heterochromatin, mechanism of satDNAs evolution as well as their functional role in centromere formation is still arguable. It would appear that centromeres are not based on satellite DNA as satellite DNA is not essential for centromere establishment, but the presence of conserved structural motifs within satDNAs, as well as functional elements such as protein binding sites or promoters indicate that there are important for centromere function. They associate with several proteins that form centromeric heterochromatin and they are proposed to drive adaptive evolution of specific centromeric histones (Ugarković, 2005).

Satellite DNAs are, commonly, considered as the most rapidly evolving part of genome as they show sequence divergence as well as changes in copy number, even between closely related species. In spite of that, some satellites like human α-satellite which has been detected in evolutionary distant species such as chicken and zebrafish are preserved for long periods exhibiting high sequence conservation (Pezer et al., 2012). Insect satDNAs show the evolutionary conservation of certain features: conserved monomer length, motifs, conserved regions, and/or secondary and tertiary structures (Palomeque and Lorite, 2008).

One of the most widespread characteristics of satDNA is an intrinsically bent structure. Clustering of A or T and regular phasing of tract of three or more A + T has been related to the degree of DNA curvature. As a result of this curvature, a superhelical tertiary is formed (Ugarković, 2005). The potential role of DNA curvature is not well established, but it is thought to be important for the tight packing of DNA and proteins in heterochromatin (Palomeque and Lorite, 2008). Palindromic sequences and dyad structures are common elements of centromeric and pericentromeric satDNAs in budding yeast, insects and humans. It is proposed that these structures could be recognized by DNA-binding proteins such as transcription factors (Pezer et al., 2012) or that they could act as nucleosome-positioning signal serving as alternative to DNA curvature (Palomeque and Lorite, 2008).

1.3. Transcription of satDNAs

Although satDNA is embedded in heterochromatin and considered transcriptionally inert, its transcription was reported in vertebrates, invertebrates and plants (Brown et al., 2012). Having in consideration simple sequence and lack of the open reading frames, transcription of satDNA was ascribed to read-through from the upstream genes and transposable elements. However, promoter elements and transcription start sites, as well as binding sites for transcription factors have been mapped with some satellites (Pezer et al., 2012).

Transcripts are heterogeneous in size and in some cases strand specific, while in other transcription proceeds from both strands. Some transcripts can be found exclusively in nucleus and others are present as polyadenylated RNA in cytoplasm.

The transcription of satDNAs is usually gender or stage specific and is often associated with differentiation and development events or certain stress conditions, particularly heat stress. In *Arabidopsis* after exposure to prolonged heat stress, heterochromatin-associated silencing is released and transcription of satDNA is significantly increased (Tittel-Elmer et al., 2010). It would appear that in *Arabidopsis* activation of transcription of repetitive elements is accompanied by heterochromatin decondesation and loss of nucleosome without loss of epigenetic marks such as DNA methylation or histone modifications (Pecinka et al., 2010). Expression of human satellite III is not detected under standard conditions but is also transiently activated by heat stress (Rizzi et al., 2004). Increasing evidence about reorganization of heterochromatin elicited by heat stress and other stress treatments could indicate possible role of satDNA in a general stress response activated in cells to cope with harmful conditions (Pezer et al., 2012).

1.4. Satellite DNAs and heterochromatin establishment

Transcripts derived from satDNAs in the form of small interfering RNAs (siRNA) participate in RNA interference-based silencing mechanism and nucleate formation of heterochromatin (Meister and Tuschl, 2004). The RNA interference-based silencing mechanism is achieved through chromatin changes, particularly characterized by histone H3-K9 methylation and/or in some particular system DNA methylation (Berstein and Allis, 2005).

The chromatin silencing is best described in fission yeast *S. pombe (*Figure 1). It is initiated by double-stranded RNA that arises from bidirectional transcription of centromeric repeats. Transcripts are further processed by the RNase III-like ribonuclease Dicer into siRNAs, which are then loaded into the RNA-induced transcriptional silencing complex (RITS) through their association with the Argonaute protein, Ago1. RITS also interacts with RNA-directed RNA polymerase complex (RDRC) which is required for the production of secondary siRNAs and sequential amplification of the silencing signal. RITS along with RDRC associates with the nascent centromeric transcript. Recruitment of RITS is achieved through the base-pairing of siRNA with nascent RNA by direct contact with RNA polymerase II elongation complex. In addition to siRNAs, the association of RITS with chromatin also requires a histone methyltransferase, Clr4 (Lippman and Martienssen, 2004). This methylation event promotes association of RITS and RDRC with target chromatin regions and further recruits heterochromatin assembly factors. It remains unclear how Clr4 is recruited to chromatin. It is proposed that association of Clr4 with heterochromatin protein Rik1 enables Clr4 binding to chromatin. It is possible that Rik1 directly links RNA interference machinery to histone H3-K9 methylation through the recognition of a specific nucleic acid substrate generated during siRNA targeting of specific chromosome domains (Verdel and Moazed, 2005).

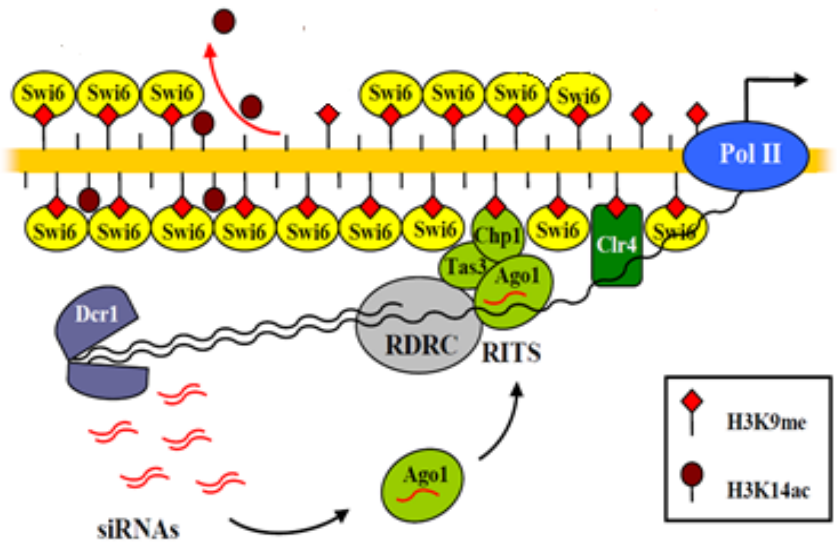


Figure 1. Mechanism of hetrochromatin formation in fission yeasr *S. pombe.* Heterochromatin formation is initiated by double-stranded RNA that arises from bidirectional transcription of centromeric repeats. Transcripts are further processed by Dicer (Dcr1) into siRNAs, which are then loaded into the RNA-induced transcriptional silencing complex (RITS) through their association with the Argonaute protein (Ago1). RITS also interacts with RNA-directed RNA polymerase complex (RDRC) which is required for the production of secondary siRNAs. RITS along with RDRC associates with the nascent centromeric transcript. Recruitment of RITS is achieved through the base-pairing of siRNA with nascent RNA by direct contact with RNA polymerase II elongation complex. In addition to siRNAs, the association of RITS with chromatin also requires a histone methyltransferase, Clr4. This methylation event promotes association of RITS and RDRC with target chromatin regions and further recruits heterochromatin assembly factors such as chromodomain protein Swi6 (taken and modified from Pezer et al., 2012).

The siRNAs cognate to satellite DNA are also involved in the epigenetic process of chromatin modification in *D. megalogaster* through H3-K9 methylation and HP1 and HP2 recruitment (Pal-Bhandra et al., 2004).

In addition to *S. pombe* and *D. megalogaster*, siRNAs derivated from satDNAs seem to be involved in the epigenetic process of chromatin modification in plants and in nematodes such as *C. elegans*. In plants, siRNAs promote heterochromatin modification by directing histone and DNA methylation. Linking RNA interference with heterochromatin silencing in mammals has been more difficult, particularly because very few repeat-derived siRNAs have been cloned. Nonetheless, the localization of methylated H3-K9 and HP1 to pericentromeric heterochromatin in mouse is abolished by RNase treatment (Lippman and Martienssen, 2004).

1.5. Disperesed TCAST satellite elements and their potential regulatory role

The existence of conserved motifs and structural properties along with increasing evidence about their transcriptional activity might indicate that in addition to participating in centromere and heterochromatin formation, satDNAs could act as *cis*-regulatory elements of gene expression as proposed by Davidson and Britten (1979). In order to perform potential regulatory function satellite elements are expected to be distributed in the vicinity of genes within euchromatic portion of the genomes.

In the red flour beetle *Tribolium castaneum* major satDNA TCAST, composed of 360 bp long tandemly arranged monomers, is localized in the centromeric and pericentromeric regions of all chromosomes comprising 35% of the whole genomic DNA (Ugarković et al., 1996; Wang et al., 2008; Feliciello et al, 2011).

TCAST satellite is transcribed in all developmental stages and its expression is temperature-sensitive indicating a putative role of satDNAs or satellite transcripts in the environmental stress response. Long primary transcripts are processed into siRNA whose increase is accompanied by the increase in H3K9me2 and H3K9me3 at heterochromatin demonstrating a putative role of TCAST siRNAs in modulating constitutive heterochromatin structure during heat stress (Pezer and Ugarković, 2012).

Whole genome sequencing project of *T. castaneum*, completed in 2008 (*Tribolium* Genome Sequencing Consortium, 2008) enabled to test the presence and distribution of satDNA repeats within euchromatin. Analysis of TCAST satellite elements distribution in the protein genes or their vicinity using that genome data set revealed association of TCAST satellite elements with 101 protein-coding gene indicating that TCAST elements might have potential role in regulation of these genes (Brajković et al., 2012).

1.6. Aims

In this study, we test existence of polymorphism of TCAST satellite elements associated with protein-coding genes in 10 different *T. castaneum* populations. We analyse TCAST elements that are associated with genes characterized as members of immunoglobulin protein superfamily in whose vicinity TCAST elements were found to be overrepresented. We also analyse polymorphism of TCAST elements inserted into introns of genes as well as the elements located very close to beginning or end of the genes (less than 2000 bp). We further analyse expression of genes that in the vicinity have polymorphic TCAST elements differing either in size or in a sequence among populations. As expression of TCAST satellite has proven to be a temperature-sensitive, indicating that it might have role in regulation of genes under stress conditions, we test the expression of the same genes under normal and heat stress conditions.

2. Materials and methods

2.1. Materials

2.1.1. Insect strains

In this study, red flour beetle *Tribolium castaneum* (Herbst, 1797) was selected as a model organism. Ten different *Tribolium* populations were included in the analysis of polymorphism of dispersed TCAST elements. Their markings and origin are listed in table 1.

Table 1. *Tribolium castaneum* strains (markings and origin)

Strain	**Origin**
GA1	wild-type strain, collected in Georgia in 1980
GA2	laboratory strain, originally used in the genome sequencing project and deriving from North American wild-type strain collected in Georgia in 1982
43	wild-type strain, collected at Kyushu Island, Japan in 1988
50	wild-type strain, collected at Indiana, USA in 2005
52	wild-type strain, collected at Bloomington, Indiana, USA in 2006
55	wild-type strain, collected at Jerez, Spain in 1991
57	wild-type strain, collected at Peru in 2002
61	wild-type strain, collected at Banos, Ecuador in 2006;
Zg Boz	wild-type strain, collected at Božjakovina, Croatia in 2010.
VT	wild-type strain, collected at Veliko Trgovišće, Croatia in 2010.

2.1.2. Solutions, reagents and other materials

Isolation of total DNA and RNA

- DNeasy Blood & Tissue kit (Qiagen)
- RNeasy Mini kit (Qiagen)

Reverse transcription

- ImProm-II Reverse Transcription System (Promega)

Polymerase chain reaction (PCR)

- DreamTaq Green PCR Master Mix (2X) (Thermo Scientific)
- Water, nuclease-free (Thermo Scientific)
- QIAquick Gel Extraction Kit

Quantitative PCR (qPCR)

- Power SYBR Green PCR Master mix (Applied Biosystems)
- Water, nuclease-free (Thermo Scientific)

Electrophoresis

-1x TAE buffer

Concentrated stock solution 50X TAE is diluted to 1X TAE working solution. This 1X solution contains in final concentrations 40mM Tris (Promega), 20mM acetic acid (Kemika), and 1mM EDTA, pH 8.0 (Promega),

- LE agarose (Roche)

- Ethidium bromide (Sigma)

- O'GeneRuler Ladder Mix (Fermentas)

-6x Orange Loading Dye (Thermo Scientific)

2.1.3. Primers for the analysis of polymorphism

Analysis of polymorphism of dispersed TCAST elements was performed by PCR using primers pairs that enabled specific amplification of TCAST satellite element located at particular site. Primers were designed using Primer3Plus software. Specificity of primers was assessed using PrimerBlast by blasting chosen primers against NCBI Genome (chromosomes from all organisms) database limiting organism field to *Tribolium castaneum*. Primer pairs marking, names and descriptions of genes in whose vicinity amplifying TCAST insert is located, sequence of primers, their length and temperature of melting (Tm) is listed in table 2.

Table 2. Primers for the analysis of polymorphism of dispersed TCAST element among *Tribolium* strains.

TCAST element	Gene name	Genbank accession number	Satellite type	Position/ distance*	Primer sequence	Primer length	Tm
2	*Probable Ser/Thr kinase*	*661947*	*satellite like*	*inside*	*F:5' GATGCACCTTGTTCGCACCTTTG 3'*	23	57,43
					R:5' AGGTCATGGCTTCCTAGCACCG 3'	22	58,38
9	*ADAM metalloprotease*	*654954*	*transposon like*	*inside*	*F:5' AAACACGTGCTAAAAGGGCTGAC 3'*	23	56,24
					R:5' CAAACGCTCGTTCAAATGCTTGC 3'	23	56,78
11	*Ser/Thr protein kinase 32B*	*655011*	*transposon like*	*inside*	*F: 5' GCTTTAACGTGCTTTAGGACAA 3'*	22	58,6
					R: 5' GCTCGAAATGAAACAGGAATAG 3'	22	58
12	*Putative uncharacterized protein*	*100141521*	*satellite like*	*inside*	*F: 5' CAGATTTCATGGAACTCATGGGC3'*	23	54,06
					R: 5'CCGGAAGAATTCAGAATACAGAAAGC3'	26	54,57
13	*Dopamine receptor 1*	*660195*	*transposon like*	*inside*	*F: 5' TGACCATTGACTATTGACTACGGA 3'*	24	53,32
					R: 5' CTGCCTGTATTACACAATTTCAACC 3'	25	53,05
17	*Transporter*	*655713*	*satellite like*	*inside*	*F: 5' CGTCCGATTTACACTCAAACTCAC 3'*	24	54,4
					R: 5' ATTAAACCGATTGAGAGAGGTTGGT 3'	25	54,24
21	*WD repeat-containing protein 47*	*657535*	*satellite like*	*inside*	*F: 5' GCTCGTCACTCCAATTAAGGTACA 3'*	24	61,5
					R: 5' CTGGTCAAAAGGCCGAATATAA 3'	22	60,6
23	*Ankyrin 2,3*	*656298*	*transposon like*	*inside*	*F: 5' CTTTGGATGGGTGTGGTAATCT3'*	22	60,1
					R: 5' AAACGTGCAGATACTGTGGTTG3'	22	60,1

TCAST element	Gene name	Genbank accession number	Satellite type	Position/ distance*	Primer sequence	Primer length	Tm
25	*Putative uncharacterized protein*	*659233*	*transposon like*	*5′/(15051 bp)*	*F:5'ATTAGTGGTGATCGCACACG3'*	20	52,68
	Coiled-coil domain containing 96	*659376*	*transposon like*	*3′/(9162 bp)*	*R:3'GCTAGATGGGAAACGTAAGAC3'*	21	50,46
27	*Putative uncharacterized protein*	*658463*	*transposon like*	*inside*	*F: 5' ATTTACTACGCGCCCAAGTTTA3'*	22	60
					R: 5' CTTCCTTAGAATTTTGCGCTGT 3'	22	60
28	*Putative uncharacterized protein*	*658191*	*transposon like*	*5′/(173881 bp)*	*F: 5' GTGTTCCTTTCGTGTCACACC 3'*	21	53,97
	Cathepsin L	*658343*	*transposon like*	*3′/(82559 bp)*	*R: 5' GCAATAATTGTGCCGAATGATACG 3'*	24	53,66
32	*Nephrin*	*664188*	*transposon like*	*inside*	*F: 5'CCGTCGCGTAATGGCTGCGA3'*	20	67
					R: 5'GCACACCCTTTGAGGTTCTGCCA3'	23	66,8
33	*Heat shock protein 70*	*100142620*	*transposon like*	*inside*	*F:5 'TGCAGATAGGAGGTGTTCTCAA3'*	22	52,85
					R:5 'GCCAAAGCCTATCCAACCTTAT3'	22	52,49
34	*N-acetylglucosaminyl-transferase VI*	*654917*	*transposon like*	*inside*	*F: 5' AGCCTAATTCGCAAGAACAGAC3'*	22	60
					R: 5' TCAGTCAGGTCAAGATCAGGTTT 3'	23	60,2
39	*Beaten path*	*100142507*	*transposon like*	*5′/(7165 bp)*	*F: 5'AATCCAAACGTCCATGCGTGTAT3'*	23	55,28
					R: 5'CAAATAAGGCATTGTAATGGCGG3'	23	53,15
40	*Putative uncharacterized protein*	*662235*	*transposon like*	*inside*	*F: 5' TCACCTGAGGACGACCACTTT3'*	21	55,08
					R: 5' GGAAGATTTGGAGAGTGATACCGAT 3'	25	54,23

TCAST element	Gene name	Genbank accession number	Satellite type	Position/ distance*	Primer sequence	Primer length	Tm
41	*Defective proboscis extension response*	*654938*	*satellite like*	*inside*	*F: 5'CTCGCGGCTGTTAAGTGGC3'*	19	62,96
					R: 5'CCGCGAAGAAATGCATGCGCT3'	21	66,29
42	*Gustatory receptor*	*662021*	*transposon like*	*3'/(17480 bp)*	*F: 5' GCTCGGCTATCCTTCTAGTTTG 3'*	22	59,5
	Ribosomal releasing factor 2	*662058*	*transposon like*	*3'/(1581 bp)*	*R: 5' AAATGGTAGCAGCGTTTCAACT 3'*	22	60,2
44	*Voltage-gated potassium channel*	*100142073*	*transposon like*	*inside*	*F: 5' AAGTGCTCGCAGTACCAAAAGT3'*	22	60,8
					R: 5'CAGAAGCTGTAATTCCCCTAGAAA3'	24	60,1
48	*PiggyBac TE*	*100142595*	*transposon like*	*inside*	*F: 5' ACTTCCTATGTTAGCCCAGTCAA3'*	23	59,2
					R: 5' CTCATCCTCCCACTTCTTCTGT3'	22	59,2
52	*Endoprotease FURIN*	*657778*	*transposon like*	*inside*	*F: 5' GCAGGAAGTCCCACAAAACTAC3'*	22	60
					R: 5' ACTAACCACCACGCATTCTTTC3'	22	60,4
58	*Putative uncharacterized protein*	*662034*	*transposon like*	*5'/(1015)*	*F: 5' CTTGCCTAGCAGCATGAATGTA 3'*	22	58,8
	Cdc73 domain protein	*657069*	*transposon like*	*5'/(27239)*	*R: 5' ATTCGGTTGAGGATGACAGTTC 3'*	22	58,4
59	*Lysine-specific demethylase 4B*	*663231*	*satellite like*	*inside*	*F: 5' GTGAGTAAGTGTGGCGTAATGTTT3'*	24	59,5
					R: 5' CAACATTCCAGGTTTCTTTCAC3'	22	58,7
67	*Ultraspiracle*	*661207*	*satellite like*	*inside*	*F: 5' TTTTGTCCGCAGGTGTACTATC 3'*	22	59,2
					R: 5' AGCTCGAAGAGGGGGAATAG 3'	20	59,8

*for those TCAST elements which were present in intergenic sequence, distance from 5' or 3' of two closest neighboring genes (one gene upstream and other downstream from insert) was listed

2. 1. 4. Primers for the analysis of gene expression

For those genes that showed polymorphism of TCAST satellite among populations, further analysis of gene expression was run using real-time PCR. Primers used in that gene expression analysis are listed in table 3. Primers were designed using Primer3plus software and their specificity assessed using PrimerBlast by blasting chosen primers against NCBI Genome (chromosomes from all organisms) database limiting organism field to *Tribolium castaneum*. Primers are designed to anneal in exon regions so that they can preferentially amplify mRNA. Positions of primers on tested genes are schematically shown in Figure 2.

The PCR product size should be kept small for accurate quantification. In this case for all three genes amplicon size is less than 220 bp (Wong and Medrano, 2005).

Table 3. Primers for the analysis of expression of genes in which TCAST polymorhism was detected

Gene	***Accession number (mRNA)***	***Primer sequence***	***Primer length***	***Tm***
2E	*XM_968077.2*	*F: 5'CCAGACCATTTCGAGGATGT 3'*	20	57,8
		R: 5'TTTATCGACGGAACGGACTC 3'	20	57,8
12E	*XM_001812009.1*	*F: 5' GGCTGCTTCGAGGATATTAAAG 3'*	22	59,4
		R: 5' CAAAGCCATACATTTTCCTTCA 3'	22	59,1
21E	*XM_963991.2*	*F: 5' GGCCAGTTGAGACGAGTTTAGA 3'*	22	60,8
		R: 5' GCATCGTAGAGTCCTTCCTGTAGT 3'	24	60,2
RPS18	*XM_962400*	*F: 5' CGAAGAGGTCGAGAAAATCG 3'*	20	57,80
		R; 5' CGTGGTCTTGGTGTGTTGAC 3'	20	59,85

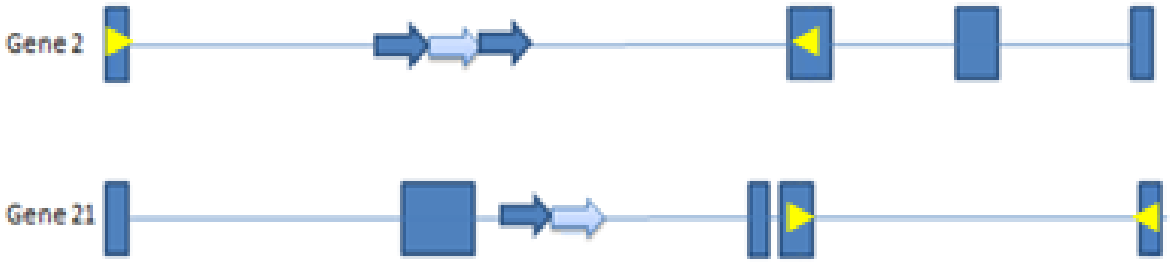


Figure 2. Structure of genes 2 and 21. TCAST elements are represented by arrows. Primers for expression, which are designed to anneal in exons, are represented by yellow lines. For gene 2 they are designed in exons 1(F) and 2 (R) and for gene 21 in exons 4(F) and 5 (R).

2.2. Methods

2. 2. 1. Insects

Cultures of *Tribolium castaneu*m were maintained on 95% whole wheat flour and 5% brewer's yeast at 25°C, no photoperiod. For analysis of gene expression after heat shock insects were exposed to 40°C overnight. After heat shock insects were allowed different time of recovery: no recovery, 10 minutes and 1h of recovery.

2. 2. 2. Isolation of total DNA using Spin-Column Protocol

DNA used in the analysis of polymorphism among *Tribolium castaneum* strains was isolated using DNeasy Blood & Tissue kit (Quiagen) according to the manufacturer's instructions. Adult beetles (approximate weight of sample 20 mg which correspond to weight of 5 insect) were homogenized in lysis buffer (buffer ALT) with addition of proteinase K, using rotor-stator homogenizator TissueRuptor (Qiagen). After thorough vortexing, tissue samples were incubated at 56°C overnight to improve lysis efficiency. Un-homogenized tissue was removed by centrifugation at 10.000 rpm for 1 min at room temperature. Supernatant was transferred to a clean tube. Before loading to DNeasy column buffer AL and 100% ethanol were added to samples and mixed well by vortexing to allow binding of DNA to the column. Samples, loaded onto spin columns, were centrifuged at 8000 rpm for 1 min. Centrifuge processing is used to remove most of the contaminants and enzyme inhibitors such as proteins and divalent cations as DNA is selectively bound to the DNeasy membrane and contaminants pass through. Remaining contaminants were removed in two sequential washing steps. DNA was eluted from the column in 200 µl buffer AE.

DNA concentration was measured using Quant-iT dsDNA HS assay kit (Invitogen) which accurately measures samples concentration from 10 pg/µl to 100 ng/µl, as average amount of DNA isolated using Dneasy kit ranged from 20 ng/µl to 60 ng/µl. After isolation samples were stored at 4°C.

2. 2. 3. PCR analysis of DNA polymorphism among *Tribolium* strains

Whole genome sequencing project of *T. castaneum*, completed in 2008 (Tribolium Genome Sequencing Consortium, 2008) enabled to test the presence and distribution of in the protein genes or in their vicinity, using that genome data set, revealed association of TCAST satellite elements with 101 protein-coding genes (Brajković et al., 2012). In analysis of distribution of TCAST elements it was determined that these elements are overrepresented near genes characterized as members of immunoglobin protein superfamily: genes associated with TCAST elements 8 (at 3'), 19 (at 5'), 25 (at 3' end), 28 (at 5'), 32 (within intron), 39 (at 5'), 40 (within intron), 41 (within intron), 62 (at 5').) Therefore, TCAST inserts in these genes were analyzed for the presence of polymorphism in different *Tribolium* populations. Additionally, for the analysis of polymorphism among populations, TCAST elements inserted in the introns or very close to the beginning or the end of a gene (distance of TCAST element from gene less than 2000 bp) were chosen. According to these criteria 38 TCAST elements were selected for the polymorphism analysis. For analysis of 24 out of these 38 elements it was possible to design primers. It was not possible to design primers for 14 elements (TCAST elements 6, 8, 19, 36, 37, 38, 43, 47, 55, 57, 60, 63, 64, 65) since genes containing enlisted elements were either poorly annotated in GenBank (unique sequence necessary for designing primers was missing) or they were enriched for repetitive elements other than TCAST element. Primers used in the analysis where designed in a unique sequence around the insert (upstream and downstream from insert) so that by sequencing of the amplicon it was possible to determine if an insert was present/absent in different population or if it varied in copy number between the populations.

PCR reactions were made using premade 2x DreamTaq Green PCR Master Mix in a final reaction volume of 20 µl. Final concentration of specific primers in reaction was 0,2 µM. In each reaction 1 µl (30-50 ng) of genomic DNA was added. Polymorphism of each insert was tested in ten *Tribolium* populations listed in table 1. No template controls were included. The following reaction conditions were used for the amplification: 94°C for 1 min, 10 cycles of 94° C for 30 s, 60°C for 30s (with annealing temperature declining in each cycle for 0, 25°C) and 72°C for 1 min followed by 20 cycle of 94°C for 30 s, 55°C for 30s and 72°C for 90s.

After amplification PCR products were visualized on 1, 2% agarose gel stained by ethidium bromide. To prepare 1, 2% agarose gel 1, 2 gram of LE agarose (Roche) was

dissolved in 100 ml TAE buffer containing ethidium bromide in final concentration 1 µg/ml. All 20 µl of reaction were loaded on the gel. To assess size of DNA molecules after electrophoresis O'GeneRuler Ladder Mix (Fermentas) was used. Electrophoresis was run at 80 V (4 V/cm) for 45 min. Gel was visualized under transiluminator using 312 nm wavelength. Photography of gel was taken and transferred to computer using Gene Snap (Syngen) program, v. 7. 05.

If PCR product size differed from expected one (insert was absent or present in different copy number), a band which contained amplicon was cut from the gel and isolated using QIAquick gel extraction kit according to the manufacturer's instruction. Gel slice was dissolved in QG buffer (3 volumes of QG was added to 1 volume of gel) at 50°C. Mixture was mixed with 1 gel volume of isopropanol and loaded onto QIAquick spin column. Contaminants were removed in one following washing step. DNA was eluted from a column in 30 µl of water and used for sequencing analysis to confirm existence of polymorphism.

2. 2. 4. Isolation of total RNA using Spin-Column protocol

RNA used in analysis of gene expression was isolated from *Tribolium* strains using RNeasy Mini kit (Qiagen) according to the manufacturer's instructions. Adult beetles (approximate weight of sample 20 mg) were rapidly frozen in liquid nitrogen to instantly stop all metabolic reactions in insects and to allow following kinetics of gene expression at different time of recovery after heat shock: no recovery, 10 minutes and 1 hour of recovery. Freezing of insects had additional purpose of aiding homogenization in the following step. Insects were homogenized in 600 µl of lysis buffer (RLT buffer) using rotor-stator homogenizator TissueRuptor. As lysis step is a limiting step in RNA isolation, to increase RNA yield maximum recommended volume of lysis buffer (600 µl) was used to ensure efficient lysis and easier homogenization. Un-homogenized tissue was removed by centrifugation at 13 500 rpm for 3 min at room temperature. Before loading to RNeasy spin column 1 volume of 70% ethanol was added to samples and mixed well by vortexing to allow binding of RNA to the column. RNA was further purified from remaining contaminants in three washing steps. After the last washing step RNeasy spin column was placed in a new collecting tube to eliminate any possible carryover of buffer RPE which may interfere with the following elution step. RNA was eluted from the column in 30 µl RNase-free water.

RNA concentration was measured using Quant-IT RNA assay kit (Invitrogen) which accurately measures RNA amount from 5 ng to 100 ng. Typically, the amount of isolated RNA exceeded 1 µg/µl. After isolation, samples were stored at -20°C.

2. 2. 5 Analysis of total RNA integrity

RNA integrity was determined using agarose gel electrophoresis. To prepare 1% agarose gel 1 gram of LE agarose (Roche) was dissolved in 100 ml TAE buffer containing ethidium bromide in final concentration of 1 µg/ml. For loading on the gel, 1 µl of sample was mixed with 4 µl of RNase-free water and 1 µl of 6x Orange Loading Dye (Thermo Scientific). To assess the size of RNA molecules after electrophoresis, O'GeneRuler Ladder Mix (Fermentas) was used. The electrophoresis was run at 40 V (2 V/cm) for 1 hour. The gel was visualized under transiluminator using 312 nm wavelength. Photography of the gel was taken and transferred to computer using Gene Snap (Syngen) program.

2. 2. 6. Reverse Transcription

RNA was reverse transcribed using ImProm-II reverse transcription system (Promega). Prior to reverse transcription reaction, 1 µl of template (approximately 1 µg of RNA) along with 1 µl of random primers (0.5µg/reaction) was incubated at 70°C for 15 min to disrupt any secondary structure that might have formed. Negative controls without reverse transcriptase were used for all samples. After incubation, a mixture was kept on ice. Reverse transcription mix was prepared according to the manufacturer's instruction. Magnesium concentration was optimized to 2 mM. Reactions were performed in a final volume of 20 µl. The following reaction conditions were used for reverse transcription: 25°C for 5 min for annealing step, 45 °C for 1 hour in extension step, 70°C for 15 min for inactivation of reverse transcriptase. Concentration of cDNA samples was determined using Quant-IT ssDNA assay kit (Invitrogen). cDNA samples were stored at -20°C.

2. 2. 7. Quantitative real-time PCR (qPCR) analysis of gene expression

cDNA samples were amplified in single tubes in an Applied Biosystem ABI7300. The qPCR reactions were done in duplicate, in 20 µl reaction volume with a specific primers concentration of 0, 5 µM and 2X Power SYBR Green PCR Master mix (Applied Biosystems). Primers nomenclature, sequences and length are listed in table 3. Amount of cDNA used in PCR reactions was approximately 30 ng. That amount of cDNA gave Cq value, which is defined as the threshold PCR cycle at which measured fluorescence is significantly higher than background levels (Wong and Medrano, 2005; Yuan et al., 2006), between 15 and 30. It was shown that Cq values in that interval can be reproducibly measured and that they are unlikely to be a major source of error when calculating ratio of expression (Karlen et al., 2007). No template controls were included. Used protocol was identical for all primer sets: 50°C for 2 min, 95°C for 1 min, 50 cycles of 95° C for 15 s, 60°C for 1 min followed by dissociation stage: 95°C for 15 s, 60°C for 1 min and 95°C for 15s. Dissociation stage is necessary to perform for constructing dissociation curve. Dissociation curve is used to determine whether during amplification primer-dimers or unspecific product were formed. Specificity of amplified product was additionally tested on agarose gel. Samples were tested in three independent runs. Data during the run were collected and processed for baseline after the run using SDS v 1.3 software (Applied Biosystems). Baseline represents level of florescence measured before any specific amplification can be detected. Baseline-corrected data were exported from SDS software and processed using LinRegPCR software v. 2012. 3 (Ruijter et al., 2009).

The calculation of starting concentrations in qPCR analysis requires an estimate of the PCR efficiency, the setting of a fluorescence threshold and determination of the Cq value. The basic equation for PCR kinetics states that the amount of amplicon after c cycles (Nc) is the starting concentration of amplicon (No) times amplification efficiency (E) to the power c (Equation 1). In this definition PCR efficiency is defined as the fold increase of the DNA amount per cycle, ranging from 1 (no amplification) to 2 (complete doubling). Equation 1 can be rearranged to calculate starting concentration (Equation 2). In this calculation as Nc value is used user-defined fluorescence threshold (Nt) located in the exponential phase of qPCR and fractional number of cycles (Cq) needed to reach that florescence threshold. Calculated No value is expressed in arbitrary fluorescence units.

$$Nc = No \times E^{c} \qquad \text{Eq. 1}$$

$$No = {}^{Nt}/_{E^{Ct}} \qquad \text{Eq. 2}$$

Analysis of qPCR data with LinRegPCR is based on the PCR efficiency that can be derived from the slope of the regression line through the data points in the exponential phase when log (fluorescence) data are plotted against cycle number (Figure 3). Points in the exponential phase used for estimation of the PCR efficiency are referred to as data points in the Window-of-Linearity (W-o-L) (Ramakers et al., 2003). It was shown that the amplicon sequence is the main contributor to the efficiency and mean efficiency values of all cDNA samples of a given amplicon shows less variation then individual efficiencies of those samples. Based on this consideration, the algorithm to set W-o-L searches for the window with the least variation between efficiencies of these reactions. The W-o-L is set to a position in which that variation is minimal. From the slope of the line plotted through 4 to 6 data points in this window, the mean efficiency is determined. For each amplicon group, a fluorescence threshold (Nt) is set at 1 cycle below the top border of the W-o-L and the Cq value is determined for the each reaction. For the calculation of No value for all samples setting of W-o-L and efficiency for that amplicon group was used (Ruijter et al., 2012).

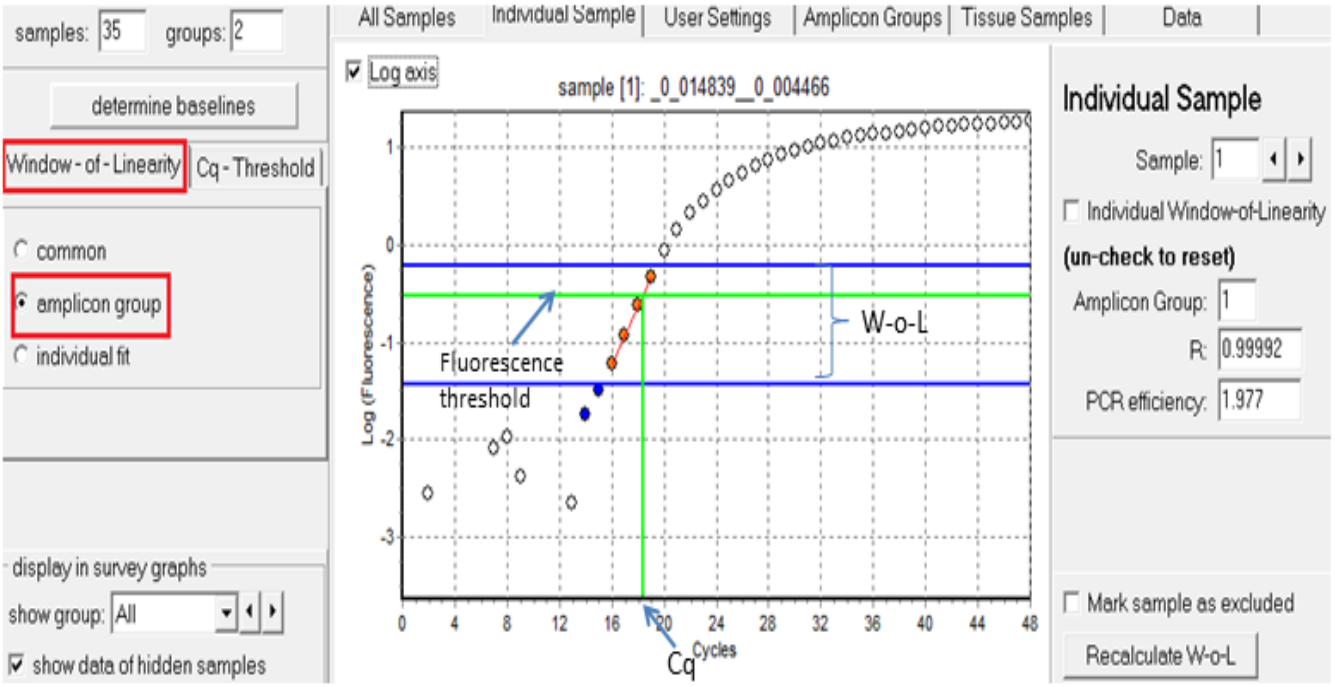


Figure 3. Screenshot of interface of LinRegPCR software. Amplification curve for one sample, determined W-o-L, fluorescence threshold and Cq for that sample are shown.

To correct for the differences in sample composition and the yield of the reverse transcription reaction, No values of tested genes were normalized using ribosomal protein S18 (RPS18). It was demonstrated that RPS18 is one of the most stable normalizers for broad scale gene expression analysis in *T. castaneum (*Toutges et al., 2010; Lord et al., 2009). To calculate the relative expression, firstly average of No value in target wells (gene of interest) of replicates per sample was taken. The same average was taken for the reference wells (endogenous control). The ratio of these two averages gives relative expression of a target gene which can then be compared between different samples.

PCR efficiency is one of the most important prerequisite in accurately quantifying specific sequence. To evaluate efficiency determined by LinRegPCR the alternative method was used (Bustin et al., 2009). This approach requires generating 5-fold serial dilution of a given sample and performing PCR reactions of each dilution. The Cq values are then plotted versus log of the dilution and a linear regression is performed. A mean efficiency can be then calculated from the slope of that line using equation 3 (Karlen et al., 2007).

$$\text{Efficiency (E)} = 10^{\left(\frac{-1}{slope}\right)} \qquad \text{Eq. 3}$$

Efficiencies determined using these two approaches: mean efficiency calculated from efficiencies linked to individual reactions as determined by LinRegPCR and efficiency that results from serial dilutions were then compared.

Preparing serial dilutions had additional function of testing for the presence of PCR inhibitors. If PCR inhibitors are present in a reaction with increasing dilution of sample, PCR efficiency should rise, as efficiency is little dependent of cDNA amount in a sample. Using LinRegPCR which uses individual efficiencies in calculating mean efficiencies such trend can be observed.

Statistical analysis of qPCR data between populations was done using GraphPad v. 6. 01. Normalized No values were compared using unpaired t test which compares the mean of two unmatched groups.

3. Results

3. 1. TCAST- related polymorphism among *Tribolium* strains

For the analysis of distribution of TCAST satellite elements in the protein genes or in their vicinity, we have chosen 38 TCAST elements. Chosen elements were either associated with immunoglobulin genes or they were located in introns of genes or in their close vicinity (distance from gene less than 2000 pb). Due to poor annotation of some genes associated with selected elements or the presence of additional repetitive elements in the vicinity of TCAST element it was not possible to analyse 14 elements. Our analysis of elements associated with immunoglobulin genes has shown no presence of polymorphism among *Tribolium* populations (data not shown). In further analysis of elements located inside or in the vicinity of genes (Table 2), we have established the existence of polymorphism in 3 elements: TCAST element 2 (associated gene: probable Ser/Thr kinase, GenBank accession number: 661947), TCAST element 12 (associated gene: putative uncharacterized protein, GenBank accession number: 100141521) and TCAST element 21 (associated gene: WD repeat containing protein, GenBank accession number: 657535). Schematic representation of genes and positions of TCAST elements within them is shown in Figure 2.

Polymorphism of element 2 reflected in different copy number of TCAST satellite unit in different populations. *Tribolium* strains GA1, 50, 61, Zg Boz, 55 and GA2 contained trimer of TCAST satellite unit arranged in tandem while strains 43, 52 and VT contained monomer of TCAST unit. TCAST element 2 in strain 57 showed presence of dimer and trimer (Figure 4). We assumed two possible hypotheses that can explain observed result for strain 57. Individuals of strain 57 are heterozygotes for element 2, or adversary, observed polymorphism of element 2 is polymorphism at individual level and as our samples contained genomic DNA from more individuals, we have detected element 2 present in separate individuals. To test our hypothesis we have amplified element 2 from genomic DNA prepared from individuals (Figure 5). As samples 1 and 2 contain trimer and samples 3 and 4 dimer we have concluded that most likely we are in the presence of individual polymorphism. Analysis of several different samples of genomic DNA of other strains did not indicate existence of individual polymorphism of element 2.

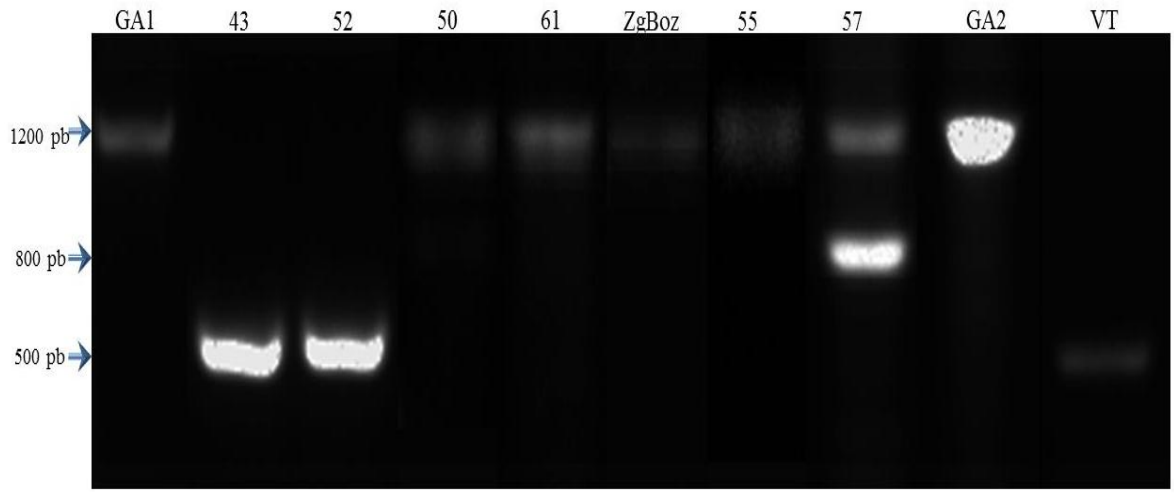


Figure 4. PCR amplicons of element 2 in *Tribolium* populations. *Tribolium* strains GA1, 50, 61, Zg Boz, 55 and GA2 contained trimer (1200 bp) of TCAST satellite unit arranged in tandem while strains 43, 52 and VT contained monomer (500 bp) of TCAST unit. TCAST element 2 in strain 57 showed presence of dimer (800 bp) and trimer (1200 bp).

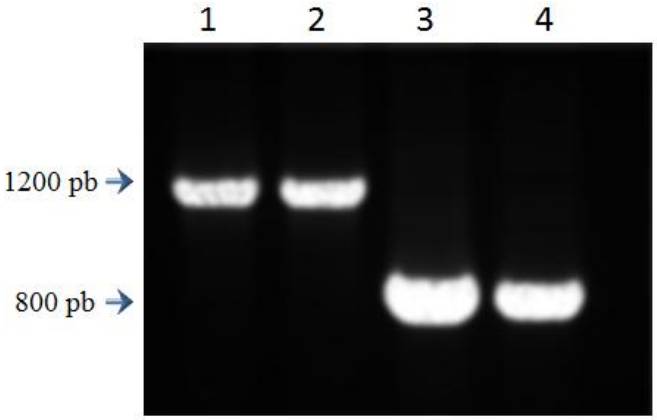


Figure 5. PCR amplicons of TCAST element 2 from genomic DNA prepared from individuals from strain 57. Samples 1 and 2 contain trimer (1200 bp) and samples 3 and 4 dimer (800 bp) which indicates that most likely polymorphism in strain 57 is at individual level.

TCAST element 12 is represented as a dimer in the genome of GA2 strain. In all other strains, except strains 43 and 52, the size of the amplicon was as expected if TCAST element 12 was present as a dimer (Figure 6). Amplicon size in strains 43 and 52 was smaller than expected (900 and 800 bp, respectively). The analysis of amplicons by sequencing showed the absence of TCAST element 12 in populations GA1 and 43 while in GA2 population TCAST insert was present, as expected. Furthermore, in strains GA1 and 43, instead of TCAST insert, another unique sequences were present. Sequences that we found in these strains share homology but they differ in size. Strain 43 has shorter form of unique sequence present in strain GA1. In strain GA2 this unique sequence was not present. Figure 7 shows sequences of TCAST element 12 from GA2 strain, sequences of amplicons from strains GA1 and 43, as well as alignment of sequences from strains GA1 and 43. BLAST analysis of unique sequence, found in GA1 and 43 strains, against repetitive elements identified in *Tribolium* gave no blast hits for query sequence. Analysis of unique sequence against *Tribolium* genome also gave no hits.

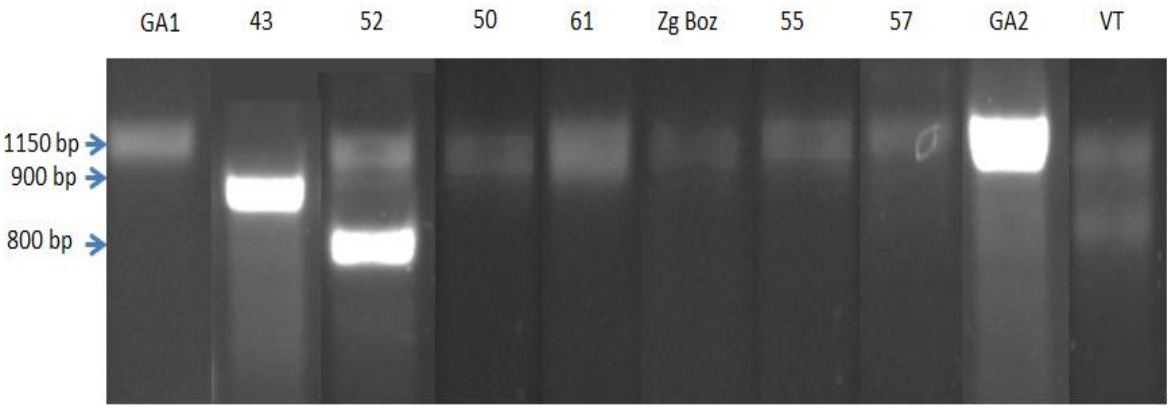


Figure 6. PCR amplicons of element 12 in *Tribolium* populations. In all strains, except strain 43 and 52 the size of the amplicon was as expected if TCAST element 12 was present. Expected size of PCR product, if TCAST element 12 was present, was 1150 bp, as it was detected in GA2 strain. Although TCAST element is missing in strains GA1 and 43, presence of another unique sequence has increased amplicon size. For that reason amplicon size of strains GA1 and GA2 is the same.

A)

>TCAST_12_strain_GA2

AACCACCATTTTATTGCAAATCTGGATCATTTTCCAGATTTCATGGAACTCATGGG
CTCATTTTTTCAGTTTTTGCTGAATAAACGCTTAGAAAAGGTAAATTTTTGGTCAT
TTTCGACCAAATTTGGTAATTTTTGTGTTCAAGTTTAGTAAAAACAGGAAAAAAT
AATGATCTTTCCTATATAAACGTGTTTAAATCCCCAGAACTTGCACAAATAATGTC
ATTTTTGATACAATTCTGTGATTTACTGGCTTACTTTTTAAGAAATTTTCGAAATA
ATTAAATTTTTGTTGAATAAATTCTTGGTCATTTTTGACCAAACTTTTTTGGTTGTA
TCATGTAATCATTTGGTTGTAAAAAAGATCTAGAACATAGTAATGGAACAAACAA
AAAAAAATGTTTGCTTCTTTAAGAAAAACCTTATGATCCTGTTTCTGTTTCCTATT
CTTATCTTTCTGGTGAGTATGTTTTTTAAATAAAGTAAATAAAATAAAAGAGGTC
AAAAAATTGTAAAAATAGCTGAGAATTTTAGACTGAAGGCTTTGTCTTTCTGTTCT
TGTGTTTTTTATTGACTGTTTTTTATTTGACAGTAAATGCTGCATTTTCGTCTTCGT
CTCAAACATTAAATACTGAATTACATTTTTGTTTAGTAACTGTTACAGTTTTCACT
TTGGTTTTTTGCTTATATCTCGAAACCGTTACTCCTATCAATTTTTATCTTTCCTTC
AAAATTAAAGCTGATAAATTTTTTTACAAAATAGTTTTTTGCATTTTTTTTGTAGG
ACCAACCATAAGCGAGTTATAGAGTTGGATATAAATAATCTTTAACAAAAATTTG
ATATAAAATAAAATGTTTGAAAAACCGAAATTAAACTAAATTAATATTAAAATAA
TTATTAATATTATTAGTATTAAAAACCGACTATAACTTTTGTATAGTAAATGTTTT
TATACCAGTTTTAAAATGAGTTTTTAATTTTGTTTGTTTTGGTTTTAAATTTGATCT
GTTGGAGGCTTTTTTTAAATGACTCGTAACTCATAACAAAAATGTGAATTTTTAGT
TTAATTTTTTTGTAAATTGAATAACAAATTTAAATTTAACTCTAAAATTTAAATTA
AAGGTGATTTAGTTAGCCCTTTGTGTTTTTTTTAGCTTTCTGTATTCTGAATTCTTC
CGGTTACGTTTTGAATTAAAAAAAAAAACGAAAGAATTATAACCAAGTTGAATC
ATGTATGTACGTAAGTAACGGATTTGCTGAAACTGAAAAACAATAAAATTCATTA
AATCAAGTAAAATTAA

>TCAST_12_strain_43

CGATTACACATACTGAATTTTGTTTAATATAGGCCTATTATTTATACCTGGGCGAT
TTTGAACTGGTGTCAAATGTGGTAATATTTTTGTTCTAGTTTAATAGAACTACGAG
ATTTAAAAACAAAAATAATGTCATTTTGACATAGTTCTGTGATTATTATAGGCTTA
CATTTTAAGAAAATTTGCGCAATATTTGAGTTTTTGCTAAATAAACGCTTAGAAA
AGGTAAATTTTTGGTCATTTTCGACCACATTTGGTAATTTTTTCGTTCAAGTTTAGT
GAAAACAGGAAGAAAAACATGATTTTTCCTCTATAAAGGTGTTTAAATCCCCAGA
ACTTGCACAAATAATGTCATTTTTGACACAATTCTGTGATTTACTGGCTTACTTTT
TAAGAAATTTTCGCAAAAATTAAGTTTTTGTTAAATAAACGCTTAGAAAAAATTT
TACCATTTTTTTTGATTGTATCATGTAATCATTTGGTTGTAAAAAACATCTTGAAC
ATAGTAATGGAACAAACAAGAAAAAAAATGTTTGCTTCTTTAAGGAAAACCTTAT
GATCCCGTTTCTGTTTCCTATTTTTATCTTTTTGGTGAGTATGTCAAAAAACTGTAA
AAATAGCTGAGAATTTTAGACTGAAGGCTTTGTCTTTCTGTTCTTGTGTTTTTTATT
GACTGTTTTTTATTTTGTTTGTTTTGGTTATATATTAAATAAAACAAAAATAAATT
TGATCTGTTGGAGGATTTTTTTAAATGACTTGTAATTCATAACAAAAATGTGAATT
TTTATTTTTATTTTTTTTTTGTAAATTATATCATAAATTTTCGAAAAACAAATTTAA
ATTTAACTCCAAAATTAAATAAGGGTGATTTAGTAG

>TCAST_12_strain_GA1

TGATTCACATACTGAATTTTGTTTAATACAGACCTATTATTTATACCTGGGCGATT
TTGAACTGGTGTCCGTAACTCGAACACGTAATAGGGGCGTTAAAGTGCTATCAAA
TGTGGTAATATTTTTGTCCTAGTTTAATAGAACTACGGGAAAAAACAGATTTTTGT
TTAAAAACAAAAATAATGTCATTTTTGACATAGTTCTGTGATTATTATAGGCTTAC
ATTTTAAGAAAATTTACGCAATACTTGAGTTTTACTGAATAAACGCTTAGAAAAG
GTAAATTTTTGGTCATTTTCGACCACATTTGGTAATTTTTGTGTTCAAGTTTAGTA
AACAGGAAGAAACCCTGATTTTTCCTATATAAACGTGTTTAAATTCCCAGAACTT
GCACAAATAATGTCATTTTTGACACAATTCTGTGATTTACTGGCTTACTTTTTAAG
AAATTTTCGAAATAATTAAATTTTTTGTTGAATAAAATTCTTGGTCATTTTTTGAC
CAAACTTTTTTGGGTTGTATCATGTAATCATTTGGTTGTAAAAACATCTTGAACAT
AGTAATGGAACAAACAAGAAAAAAAAATGTTTGCTTCTTTAAGAAAAACCTTAT
GATCCTGTTTCTGTTTCGTATCCTTATCTTTCTGGTGAGTATGTTTTTTAAATAAAG
TAAATAAAATAAAAGAGGTCAAAAAATTGTAAAAATAGCTGAAAATTTTAGACT
GAAGGCTTTGTCTTTCTGTTCTTGTGTTTTTTATTGACTGTTTTTTATTTTTACATTA
TTATTTACATTATTTTACATTTCTATTATTATTGTTTGTTTCGGTTTTAAATTTCATC
TGTTGGAGGCTTTTTTTTAAATGACTCGTAACTCATAACAAAAATGTGAATCTTTA
ATTTAATTTTTTGTAAATTATATCATAAATTTTGAATAACGAGAATTAAATTAACT
CTAAAATTAAATAAAGGTGATTTTAGTAGCCCTTGTGTTTTA

B)

C)

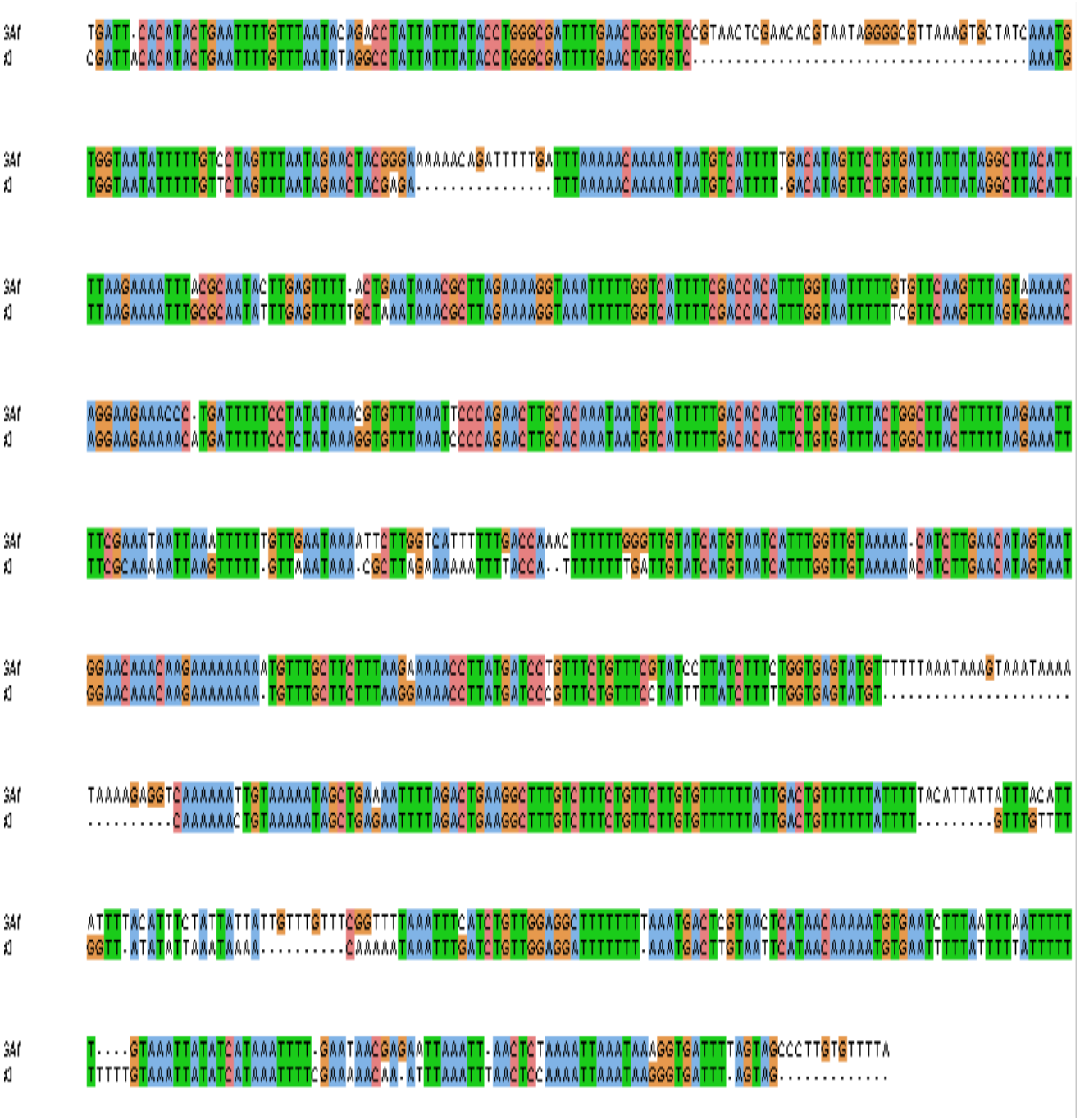


Figure 7. A) Sequences of TCAST element 12 with flanking regions in strains GA2, 43 and GA1. B) Shematical representation of TCAST element 12 and its flanking regions. Parts of sequences shared among all three strains are colored green. Unique sequence present only in strains GA1 and 61 is colored orange. Unique sequence present only in GA2 strain is colored red. TCAST element is represented with arrows. C) Alignment between unique sequences present im strains GA1 and 43. Alignment shows that sequences in these strains share homology but differ in size.

Amplification of TCAST element 21 from genomic DNA of *Tribolium* strains gave more than one product in strains 50, 52, Zg Boz, 55, 57 and VT while in strains GA1, 61, 43 and GA2 only one PCR product was present. For further analysis we chose strains GA1, 61 and GA2.
In strains GA1, 61 and GA2 analysis of TCAST element 21 showed similar organization as in element 12. TCAST element was present in strain GA2 but was absent in strains GA1 and 61 (Figure 8). In strains GA1 and 61 another unique sequences were present which again shared homology but differ in size, meaning that sequence present in strain 61 is a shorter form of sequence present in GA1. Figure 9 shows sequences of TCAST element 21 from GA2 strain, sequences of amplicons from strains GA1 and 61, as well as alignment of unique sequences from strains GA1 and 61. BLAST analysis of unique sequences present in GA1 and 61 strains could not identify existence of any homology with repetitive elements present in *Tribolium.*

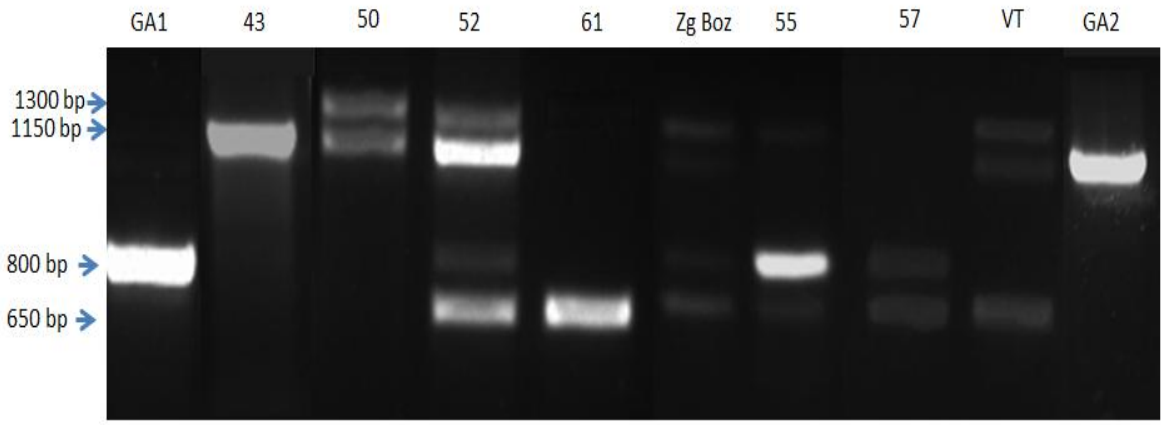


Figure 8. PCR amplicons of element 21 in *Tribolium* populations. Amplification of TCAST element 21 from genomic DNA of *Tribolium* strains gave more then one product in strains 50, 52, Zg Boz, 55, 57 and VT while in strains GA1, 61, 43 and GA2 only one PCR product was present. For further analysis we chose strains GA1, 61 and GA2. Expected size of PCR product, if TCAST element 21 was present, was 1180 bp, as it was detected in GA2 strain. In strains GA1 and 61, instead of TCAST insert, another unique sequences were present.

A)

>TCAST_21_strain_GA1

GTATTTTTATGTTCAGTAAAACATGAATGCTCAGTACAGCACTGTCAGTTTCTTATACTAATAAAGA
CGCAGTTCAGTGTTACCTGAACAATTGTGAATTATTTATAATAAATATGATAAATCTGTACCTATTC
TGACCACAACCATCTCCAATAATTTTTTTTGTAGAAGACATCAAAGCACGATAAAATTTTACTTTTT
AACTTAAAAAACAACTGTTTAAATTTTATTTAATATTTTTGAACAATTTGAAACTCGCTACATCTAT
TTTAATGTTGGATTAAAATATACGCCTATCGTCGATTTAAAAAAAACTGGTTAAAAGTTTAAATTCA
AAGTTAAAGTTAAACTAATTAACAGCAATCAAATGATTCAGATCACGTGACCAAATTTGACCAATC
ACAGTGTTAAAAGCAAGGGCTAAATAACAGCGGATTAACTTTTGATTCGTTTTCAGTGGTCACTCG
ATTTGTAGTGCAAGAAAAAAATTTGATTGCAAATTGCAATTCATGTTCTGTATTCGGTTCTACGAAA
AAGTGGTCTTATCAGGTAAAAATGTGTTATCATCAAAAGTGTAACCAAAACGAAATATGAATTTTT
TATGATATGAGATGAAGAGGAATGATTTACAACTTTTTATATTACTGTATTTATTCTGTTACTCGTT
CAAAAAAACCGGGCCAACCACAGTCGAAA

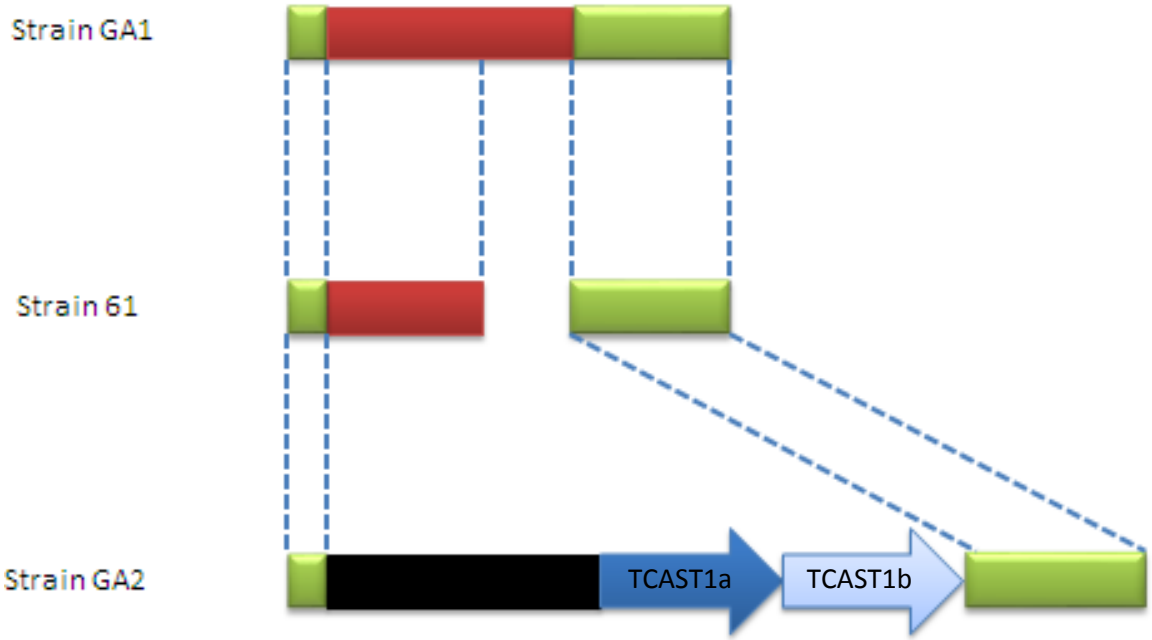
>TCAST_21_strain_61

GTATTTTTATGTTCAGTAAACATGAATGCTCAGTACAAGCACTGCGTTTCTTATACTAATAAAGACG
CAGTTCAGTGTTACCTGAACAATTGTGAATTATTTTTTATTAAAATATGATAAATCTGTACCTATTC
TGACCACAACCATCTCCAATAATTTTTTTTTTTGTAGTAGACATCAAAGCACGATAAAATCTTACTTT
TTTAACTTAAATGACAAATTTTATTTAATCTTTTTGATCAATTTGAAACTCGCTACATCTATTTTAAT
GTTGGATTATAATATACGCCTATCGTCGATTTTAAAAAACTGGTGAAAAGTTTAAATTCAAAGTAA
AAGTTGATTGCAAATTGCAATTCATGTTTGTATTCGTTTCTACGAAAAAGTGGTCTTATCAGGTAAA
AATGTGTTATCATCAAAAGTGTAACCAAAACGAAATGTGATTTTTTTATGATGAGATGTTTGAGGA
ATGATTTACAACTTTTTATATTACTGTATTATCTGAACTCGTACAAGAAA

>TCAST_21_strain_GA2

ATATGCGCAAATTTAGCTACACGATCCTCAAGCACAAGTACATAGAGCTGTTATGTATCAAGTCGG
AGGCCAACGTGCAAGGGAACAGCGTCGACAACGCGGTGGAAGAAGTCGTTAAAGTCCTAAACGAT
TTGGAAAAGTTTGCGCCTTCCAAGGAGGATTACTCGAATCTGTGCTTACTTCTGACGCTGCCGCGGC
TCACCGATCATTTGCAATACAAAGACTGGAATCCGAGTAACGCCCGAGTGCAGTGCTTTCGAGAGG
TGTATCCTCTAGTAGCTGAATTTCTACCAGGAGATAGGAAATCGACAGATGCCAATGCCAATAACA
GCGCAAAAAACGACAGACTGATCCAATTAATTATCAAAGGCATTCTCTACGAATCCTGTGTGAATT
ACTGTCAAGCCAAAGCGACAGGAACCACCGAATCTCAATCGCACGAAATGAATTTCTCGAGATTAC
TTGACGGCAGTGTTGGTTTTAGTGACTCAGATCTTAGTTTATTGTCTTGGTTGCAAAGCGTTCCTTC
GGAAACGTTCTCTGTACCCTTCGAGCAACGAACCCTCAACGTGGACGTGGAACGTCTGGAAAGACC
TTCCTTGGAGACTTCCTGGACCGAGCACATGCTCGTCACTCCAATTAAGGTACATGTTACAGTACA
GTATTTTTATGTTCAGTGAAAACAATGAATGCTCAATAAACTCAACTGAGTTTCTTATACTAAACTT
TATACTATTAAACACAACTGTTTAATATTTCATTTAATATTTTTGAACAATTTGAAACTCACTTCTAT
TTTAATGTTGGATTTAAATATACACCTATCGTCGATTTTAAAAAACTGGTTAAAAATTTTAATTCAA
AGTAAAAGTTAAGCTAATTAACAGCGGATTAACTTTTGATTCGTCTTTAGTGTCACTCAAATTTTAT
TGCAAGAAAAAAAAATTGTTCGCAAATTGATAAAAATTTATAGGAGTAACTGTTTTCGAGATATAA
GCAAAAAACCAAAAAACTGTTACCTTAATTAAGCGCTTGCATATAAGCAAATTTAGTACTATTAGT
ATTATTATTATTAATGTTATTGTTATTAATATGACGACTAAAGACGACTCTGAAGACTTTAGTGACT
TTTATGTCTCCTCCTTTTTCCTCTACTAAATTGTTAGACACAATTAATTTAGTTTAATTTCAGTTTTTC
AAACACTTTATTTTAAAGCAAATTTTTGTTAAAGATTATTTACATCCAACTCTATAACTCGCTTATG
GTTGGTCCTACAAGAAAATTGCAAATAACTATTTTGTAGGAAATTTTATCAGCTTTAATTTTGAAAA
AAAGATAAAAATTGAGAGGAGTAACTGTTTTCGAGATATAAGCAAAAAACCAAAATGAAAACTGT
AACTGTTACTGAACAAAAATGTAATTCAGTTTTCAATGTTTGAGACGAAGACGAAAATGCAGCATT
TAAAAACGACTCTGAAGACTTCAGTGACTCCTTTTTCATCTAATAAAGTGTTGGACCCAATAAATTA
AGTGTATTTTAAATCACAAGACATTTTCTGTTTTCTGTTTTCTATTTCTAAAATTTGATTGCAAATTG
CAGTTCACGTTCTGTATTCGGTTCTACGAAAAAGTGGTCTTATCAGGTAAAAATGTGTTATCATCAA
AAGTGTAACCAAAACGAAATGTGAATTTTTTATGATATGAGATGTTTAGGAATGATTTTACAACTTT
TTATATTACTGTATTTATTCTGTACTCGTTTCAAAAAA

B)



C)

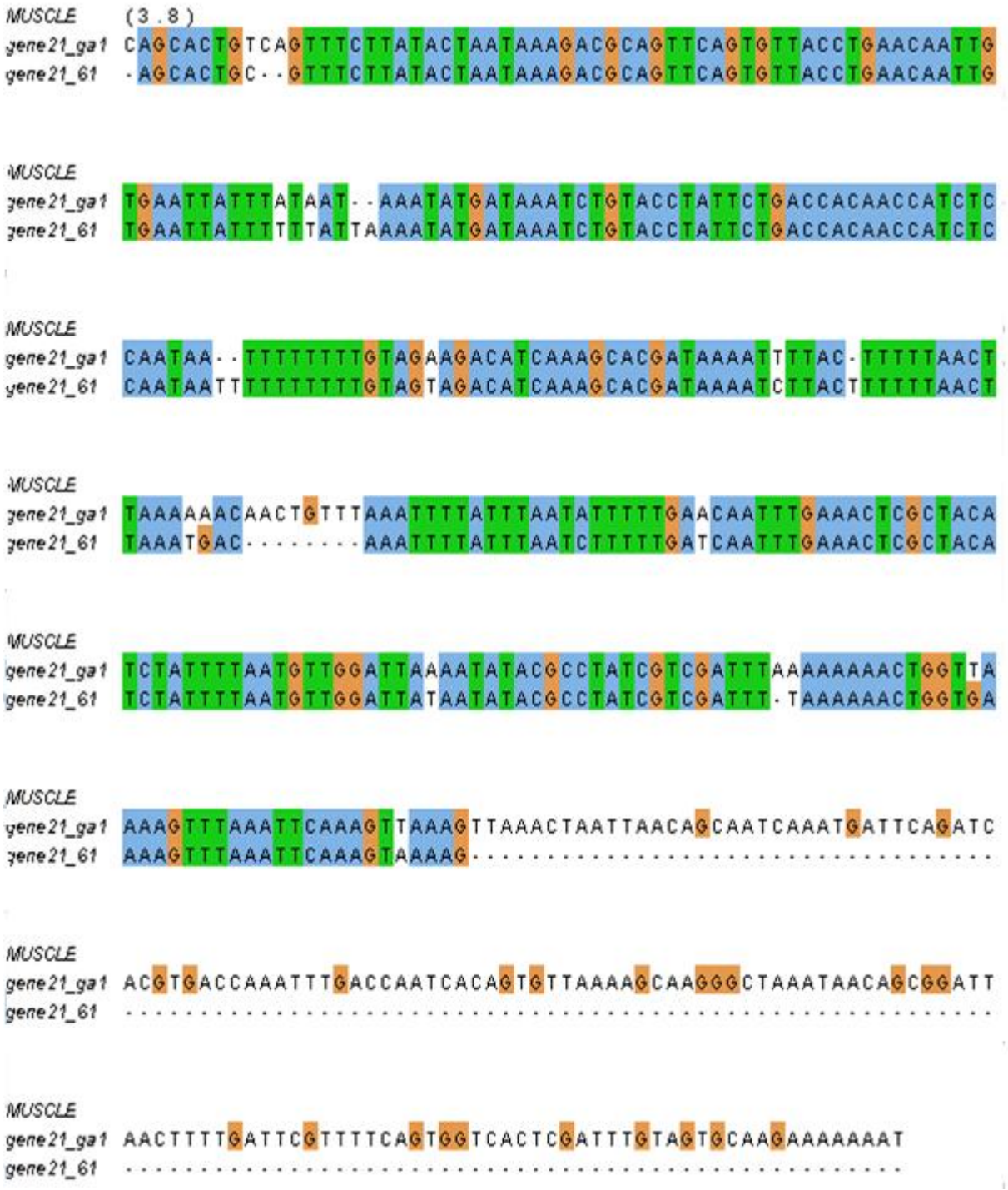


Figure 9. A) Sequences of TCAST element 21 with flanking regions in strains GA1, 61 and GA2. B) Shematical representation of TCAST element 21 and its flanking regions. Part of sequences shared among all three strains are colored green, Unique sequence present only in strains GA1 and 61 is colored in red. Unique sequence present only in GA2 strain is colored black. TCAST element is represented with arrows. C) Alignment between unique sequences present im strains GA1 and 61. Alignment shows that sequences in these strains share homology but differ in size.

3. 2. Structure of TCAST element 2

Sequence analysis of TCAST element 2 in strains GA1, 50, 61, Zg Boz, 55 and GA2 has shown that the element is composed of three copies of the satellite DNA unit repeat, arranged in a tandem. TCAST satDNA in pericentromeric region is composed of two types of monomers, TCAST 1a and TCAST 1b, which are mutually interspersed. Two monomers share average homology of 79% and are of similar sizes: 362 bp and 377 bp respectively (Brajković et al., 2012).

Dispersed TCAST element 2, when represented by a trimer, has a similar interspersed structure as the one characteristic for TCAST satellite within heterochromatin (TCAST 1a-TCAST 1b-TCAST 1a) (Figure 10a). In VT, 43 and 52 strains, where element 2 is present as a monomer, TCAST insert shows clear homology with a second TCAST1a monomer present within element 2 trimer (Figure 10b). Organization of TCAST element 2, either in a form of trimer or monomer, is schematically shown on Figure 10c. Inserted TCAST element 2 is characterized by the presence of a short direct duplicated region at both ends in all population tested, suggesting the mechanism of dispersion is based probably on site specific recombination (see also discussion).

A)

AGTTGAAAAAAAATATGTTTGGGAAAAAATGAATTTTGTATTGAAAATTATCCGATGCACCTTGTTCGCACCTT
TGTAACAATTAGTTATCTTGACAGATTAAATGATAACATTTAATTGAACA**AATGATAA**AATTTCCTACAAAAT
AGTTATTTGCATTTTTCTTGTAGGACC**AACCATAAGCGAGTTATAGAGTTGAATATAAATAATCTTTAACAAA**
AATTTGCTCTAAAATAAAATGTTTGAAAAACTGAAATTAAACTAAATTAATTGTGTCTAACACTTTGGTAGA
GGAAAAAGAAAAAGACATAAAAGTCACTAAAGTCTTAAGAATCGTCTTTAGTCGTCATATTATTAACAATAA
CATTAATAATAATAATACTAATAGTACTAAATTTGCTTATATGCAAGCGCTTAATTAAGGTAACAGTTTTTTG
GTTTTTTGCTTATATCTCGAAAACAGTTACTCCTATCAATTTTTATCTTTTTACCAAAATTAAAGCTGATAAAAT
TTCCTACGAAATAGTTATTTGCATTTTTCATGTAGGACTAACCATAAGCGAGATATAAGTTTGAAAATAATTA
ATAATTAAAAAAAAATGTGCTTTAACAGAAAATGTCTTGTGATTTAAAGTACACTAAAATTATTGGGTCCAA
CACATTATTAGAGGAATAAGGAGGAGACATAAAAGTCACTGAAGTCTTCAAAGTCGTTTTTAAATGCTGCAT
TTTCGTCTTCGTCTCAAACATTAAAAACTGAATTACATTTTTGTTCAGTAACAGTTACAGTTTTCATTTTGGTTT
TTTGCTTATATCTCGAAAGCAGTTACTCCTATCAATTTTTATCTTTCTTTCAAAATTAA**AGCTGATAAAATTTCC**
TACAAAATAGTTATTTGCATTTTTCTTGTAGGACTAACCATAAGCTAGTTATAGAGTTGGATATAAATAATCT
TTAACAAAAATTTGCTCTAAAATAAAATGTTTGAAAAACTGAAATTACACTAAATTAATTGGGTCTAACACTT
TAGTAGAGGAAAAAGGAGGAGACATAAAAGTCGCTAAAGTCTTCAGAGTCGTCTTTAGTCGTCATATTAAT
AACAATAACATTAATAATAATAATACTAATAGTATTAAATTTGCTTATATGCAAGCGCTTAATTAAGGTAAT
GATAACATTTGTTCTTAAGTAACACAGTATTTTTGGCGGGCGGTAAAACTTGCAGTACCGCACGTCCACACGT
GTATGTCATTTGTTATCGGTGCTAGGAAGCCATGACCTATTACCACTGCTATTATTTCTTCATTAGGAAGCTTGC
AGTGATTAAATTATCAAAAAAAAATGACACATTTTCCTAATCTCTTC

B)

ATCTTGACAGATTAAATGATACATTTATTGACAATGATAAAATTTCCTACAAAATATTGTTATTTGCATTTT
TCTTGTAGGACCAACCATAAGCGAGTTATACAGTTGGATATAAATAATCTTTAACAAAAATTTGCTTTAAAAT
AAAATGTTTGAAAAACTGAAATTACACTAAATTAATTGGGTCTAACACTTTAGTAGAGGAAAAAGGAGGAG
ACATAAAAGTCGCTAAAGTCTTCAGAGTCGTCTTTAGTCGTCATATTAATAACAATAACATTAATAATAATAA
TACTAATAGTATTAAATTTGCTTATATGCAAGCGCTTAATTAAGGTAATGATAACATTTGTTCTTAAGTAAC
ACAGTATTTTTGGCGGGCGGTAAAACTTGCAGTACCGCACGTCCACACGTGTATGTCATTTGTTATCGGTGCT
AGGAAGCCATGACCTT

C)

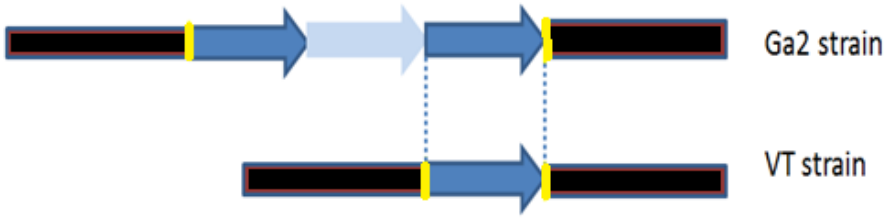


Figure 10. Structure of TCAST element 2. The sequence of the two flanking regions (indicated in black) and of three contiguous TCAST satDNA repeats of element 2 in GA2 strain (TCAST 1a is indicated in dark blue and TCAST 1b is indicated in light blue) and sequence of a single TCAST repeat in VT strain, are shown on A and B, respectively. Direct duplicated regions at both ends of the TCAST1 are underlined and enlarged. Figure 10c is a schematic illustration of the different organization of element 2 in GA2 and VT strains. In VT strain, where element is present in one copy, TCAST insert shows clear homology with a second TCAST1a present in elements 2 trimer . Unique flanking region of intron where element 2 is inserted is indicated in black. TCAST 1a monomer is indicated in dark blue and TCAST 1b in light blue. Direct duplicated regions are indicated by yellow lines.

3. 3. Expression of genes associated with polymorphic TCAST elements

To test the potential role of dispersed TCAST elements in the regulation of nearby genes we have analysed expression of genes associated with TCAST elements 2, 12 and 21, that had shown the presence of polymorphism among *T. castaneum* strains. Purpose of the analysis was to determine if there is any difference in expression of genes with respect to the presence or absence of TCAST elements or with respect to different copy number of elements present at one locus. For simplification, we have named analysed genes by number of TCAST element that they contain. Gene 2 served as a model to test if difference in copy number of dispersed TCAST element can influence gene expression in physiological conditions. To confirm that individual insects, from which RNA used in expression analysis was obtained, in fact carry assigned polymorphism, we have amplified element 2 from genomic DNA which was present in a small amount in RNA samples (Figure 11). Although, preliminary analysis of polymorphism of element 2 did not indicate existence of individual polymorphism among strains other than strain 57, additional analysis of RNA samples has shown that in other strains individual polymorphism is present, as well. Individual polymorphism in other strains is probably present in lower frequency than in strain 57. Element 2, in strains 50 and Zg Boz was originally determined to be a trimer, but analysis of RNA samples has shown that in these samples that element is a monomer. In contrast, VT strain which was originally assigned as a monomer has amplified as a trimer.

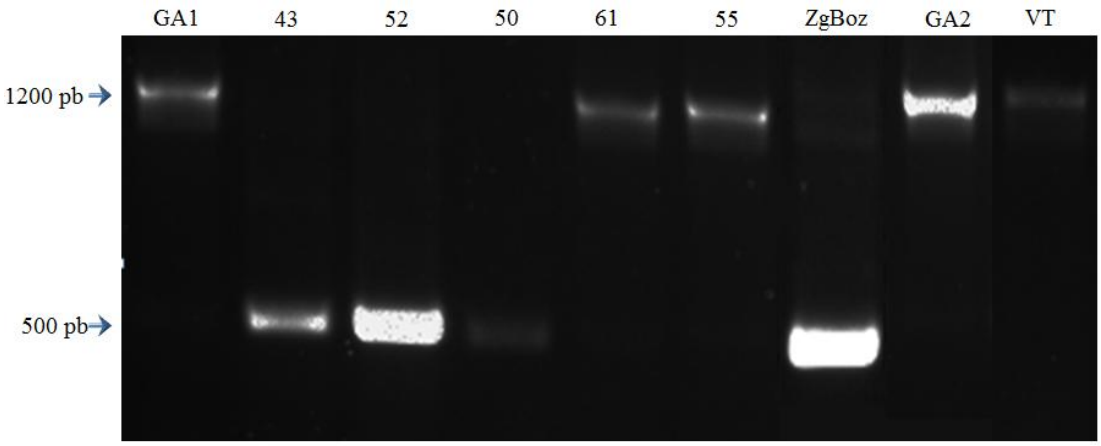


Figure 11. PCR amplicon of TCAST element 2 from genomic DNA contained in RNA sample. Preliminary analysis of polymorphism of element 2 did not indicate existence of individual polymorphism among strains other than strain 57. Additional analysis of RNA samples has shown that in other strains individual polymorphism is present, as well. Element 2, in strains 50 and Zg Boz was originally determined to be a trimer but analysis of RNA samples has shown that in these samples that element is a monomer. In contrast, VT strain which was originally assigned as a monomer has amplified as a trimer.

Comparison of expression of populations that contain element 2 in the form of a monomer and in the form of a trimer was done in accordance to the results of amplification that we obtained from the analysis of RNA samples (Figure 12). Analysis of gene 2 expression by qPCR in populations that contain trimers (GA1, 61, 55, GA2 and VT) with respect to the populations that contain monomers (43, 52, 50, Zg Boz) has shown no significant difference between these two groups. Mean No value of gene 2 in strains that contained trimers was 0, 2345 ± 0, 03197 and mean No value of other group was 0, 2655 ± 0, 01703. P- value determined by un-paired t-test was 0, 403 with 95 % confidence interval being -0, 04461 to 0, 1065.

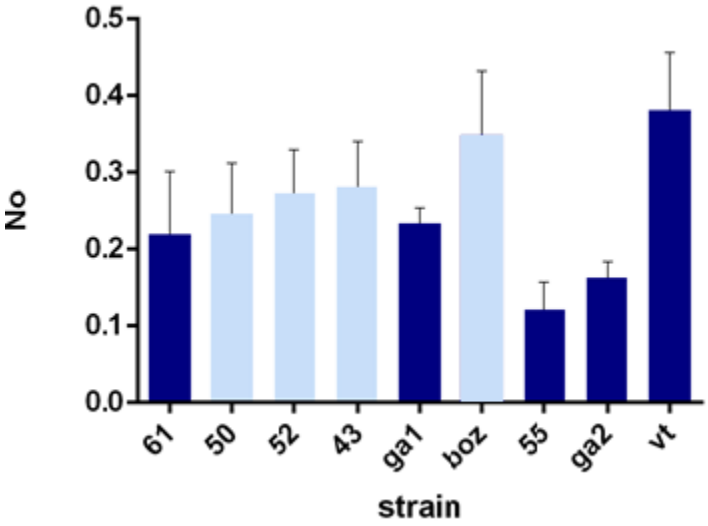


Figure 12. Expression analysis of gene 2 among *Tribolium* populations. Analysis of No values of gene 2, obtained by LinRegPCR method from qPCR data did not show any significant difference between gene expression in different strains. Strains that contain trimers are indicated by dark blue color and strain that contain monomers by light blue. The data shown are the mean and standard deviation of duplicate PCRs resulting from three independent experiments.

Genes 12 and 21 where used as model to establish if presence of TCAST element can affect gene expression. Unfortunately gene 12 was expressed in a very low amount with Cq value above 35 and for that reason was eliminated from further analysis. Comparison of expression of gene 21 between populations GA1 and 61 showed no significant difference between populations (Figure 13). TCAST element 21 was absent in 61 population and present in GA2 population. Mean No value of gene 21 in GA2 population where insert was present was 0, 1477 ± 0, 02835 and mean No value of other group was 0, 2460 ± 0, 02835. P- value determined by un-paired t-test was 0, 2671 with 95 % confidence interval being -0, 3052 to 0, 1085.

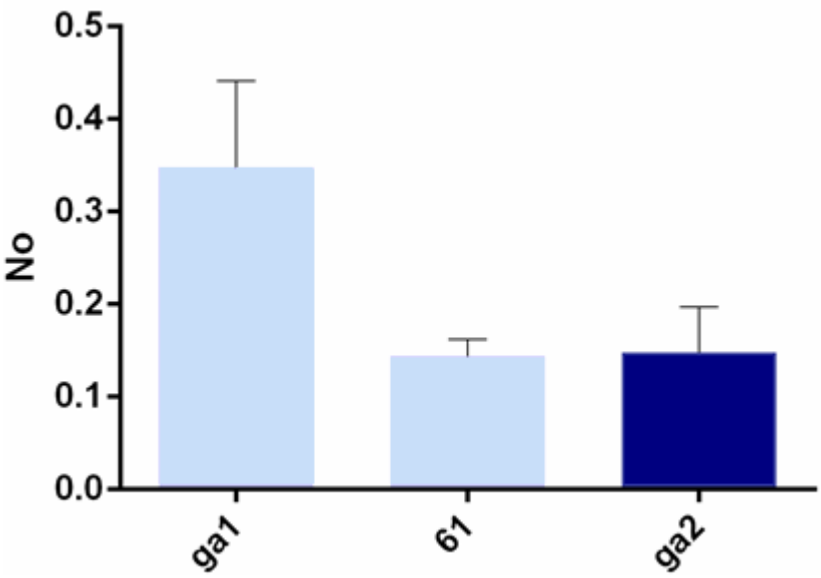


Figure 13. Expression analysis of gene 21 among *Tribolium* populations GA1, 61, GA2. Analysis of No values of gene 21, obtained by LinRegPCR method from qPCR data did not show any significant difference between gene expression in different strains. Strain GA2 contains TCAST insert and it is indicated by dark blue color, while strains GA1 and 61 do not contain insert, so are indicated in light blue. The data shown are the mean and standard deviation of duplicate PCRs resulting from three independent experiments.

Considering evidence of increased TCAST satellite transcription as well as presence of TCAST related siRNA after heat stress (Pezer and Ugarković, 2012), which could indicate role of TCAST in regulation of genes under stress conditions, we have tested expression of genes 2 (Figure 14) and 21 (Figure 15) after overnight (12 h) exposure of adult beetles of GA2 and 61 strains to temperature stress at 42°C. Transcription of gene 2 after heat stress was downregulated in GA2 strain. After heat stress gene 2 was expressed 3.2 times less in comparison with control which was not subjected to heat stress. Considerable difference in transcription of gene 2 was observed after 1 h of recovery at 25°C. Expression of gene 2 in GA2 strain was 14.8 times lower with respect to its expression under normal conditions (Figure 14).

Gene 21 was downregulated after heat stress in strains 61 and GA2. When compared to control, which was not subjected to heat shock, transcription of gene 21 was decreased 2.4 times for GA2 strain and 1.92 times for strain 61. Significant difference in expression of gene 21 between strain 61 and GA2 was observed after 1 h of recovery at 25°C. Expression of gene 21 in GA2 strain, which contained TCAST element was 3.1 times lower after 1 h of recovery period compared to the expression of the control. However, in strain 61 which does not have

TCAST element inserted within a gene, expression of gene 21 after 1h of recovery from heat stress has returned to the level of the control (Figure 15). This result shows significantly slower recovery of expression of gene 21 with TCAST element within intron after heat stress relative to the expression of gene 21 which lacks TCAST element. This indicates influence of TCAST element on the expression of gene under specific conditions following heat stress.

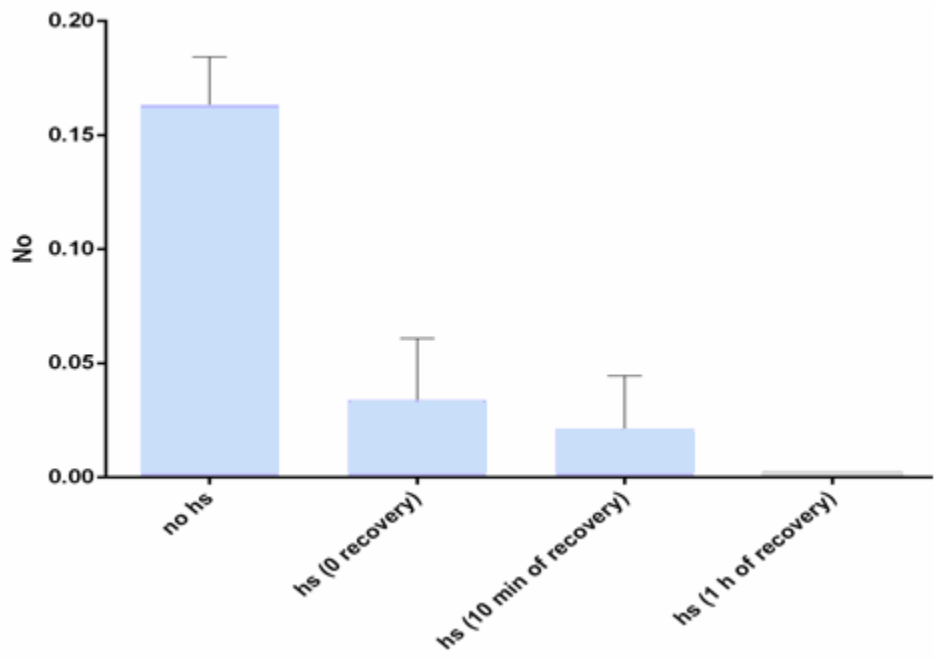


Figure 14. Expression analysis of gene 2 in GA2 population after heat stress. After heat stress (hs-0 recovery) transcription of gene 2 was downregulated. After one hour of recovery significant downregulation of gene 2 with respect to the control was observed. The data shown are the mean and standard deviation of duplicate PCRs resulting from two independent experiments.

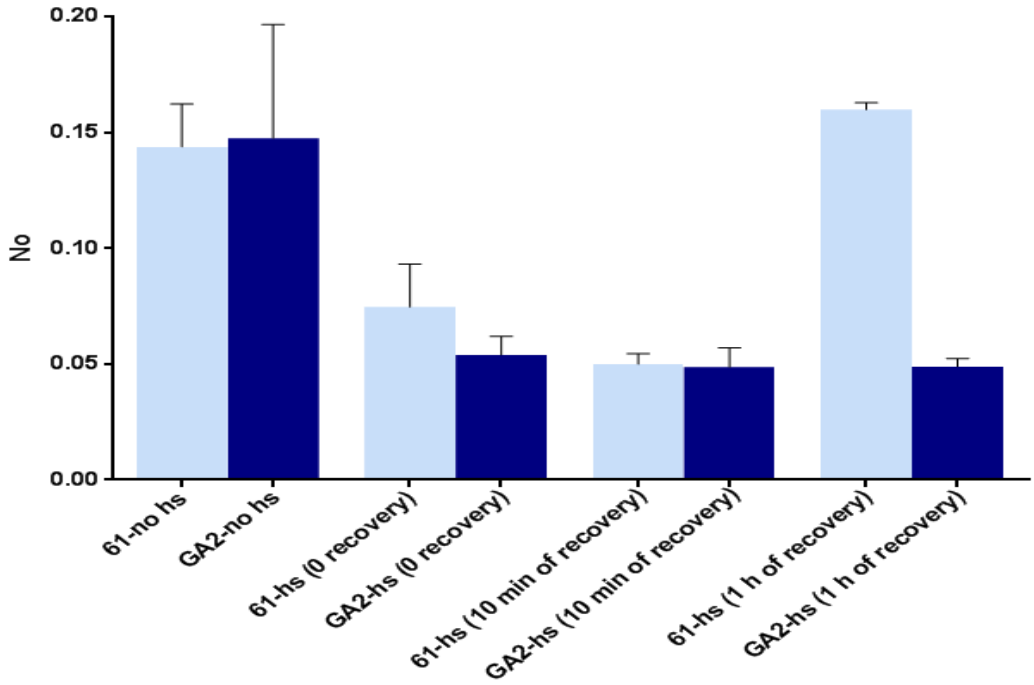


Figure 15. Expression analysis of gene 21 among *Tribolium* populations GA2 and 61 after heat stress. After heat stress (hs-0 recovery) transcription of gene 21 in both strains was downregulated. After one hour of recovery significant downregulation of gene 21 was observed in GA2 strain which contained TCAST insert, while in strain 61 where TCAST element was absent the level of expression has returned to normal. The data shown are the mean and standard deviation of duplicate PCRs resulting from two independent experiments.

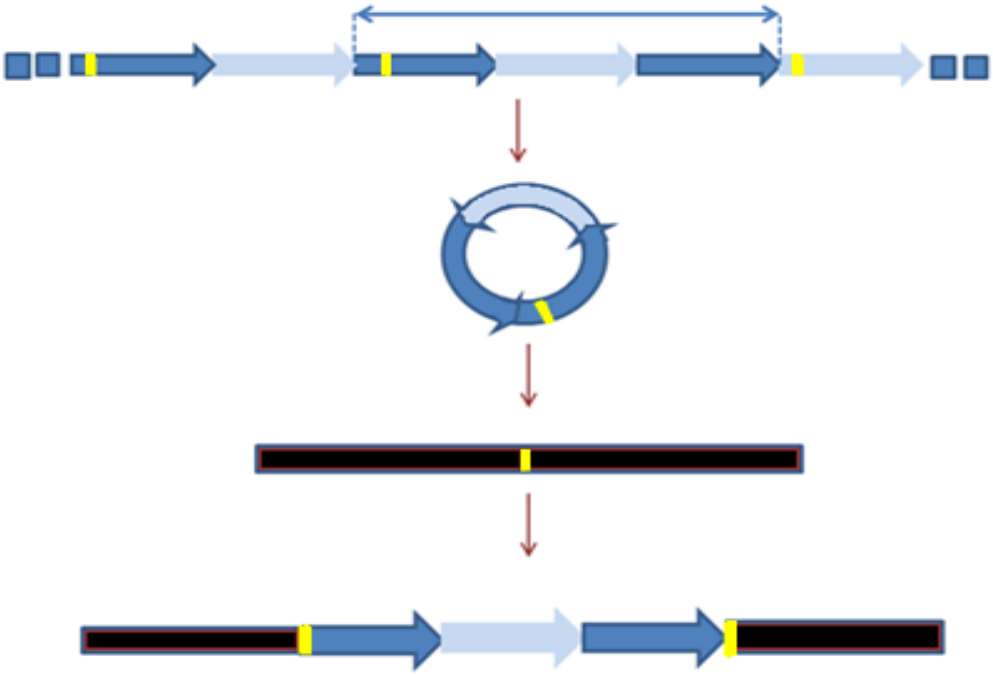
4. Discussion

Our analysis of TCAST related polymorphism among *Tribolium* strains has shown the presence of polymorphism of 3 TCAST elements: 2, 12 and 21. The expression of gene 12 in normal condition is very low so the presence or absence of TCAST element in that gene, even if dispersed TCAST elements are important regulators of gene activity, probably would not influence expression of that gene. The polymorphism of element 2 is related to change in copy number of TCAST monomer. In our study we have not observed that the difference in copy number of dispersed satellite-like element has any effect on gene regulation which might indicate that for that role only the presence of satellite element is necessary. In that case, polymorphism that could have some influence on gene regulation was observed only in connection to one element, element 21. Observed rare frequency of polymorphisms among populations would be in accordance with assumption that TCAST elements serve as regulatory elements.

TCAST satellite is a major satellite that encompasses centromeric and pericentromeric regions of all 20 chromosomes and it comprises 35-40% of the whole genome (Ugarković et al., 1996). In this study we consider the possibility that TCAST element dispersed in genome of insect *Tribolium castaneum* can regulate expression of gene in its vicinity. When addressing that issue, one of the important questions that arises is the mechanism which could allow spreading of TCAST satellite-like elements through the genome. Structure of insert 2 present in the form of a monomer, dimer and a trimer provides bases for the possible model of dispersion of satellite-like elements (Figure 16). This model postulates that a circular DNA containing three copies of TCAST1 repeat is excised from a satDNA array by intramolecular homologous recombination. The site specific insertion of these circularized repeat at a homologous chromosomal site generates a structure identical to that identified in this study. The presence of extrachromosomal circular DNAs (eccDNAs) formed by satDNA tandem arrays has been reported in many organisms (Pont et al., 1987; Cohen et al., 1997; Cohen et al., 1999). Feliciello et al. (2005, 2006) proposed that satDNA evolution occur by a DNA repair mechanism most probably based on rolling circle amplification. The extrachromosomal circular DNAs formed by satDNA might be a direct consequence of this mechanism or they could be formed by intramolecular recombination events. Other possible model includes

unequal crossing over, slippage replication, transposition, or RNA-mediated exchange (Palomeque and Lorite, 2008).

A)



B)

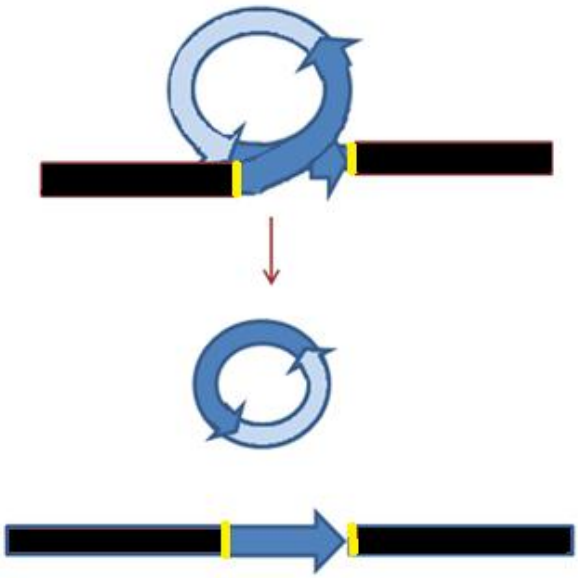


Figure 16. A) Model of the generation of dispersed satellite-like elements. This model postulates that a circular DNA containing e.g. three copies of TCAST1 repeat is excised from a satDNA array by intramolecular homologous recombination. The site specific insertion of these circularized repeat at a homologous chromosomal site generates a structure identical to that identified in this study. Unique region of intron where element 2 is inserted is indicated in black. TCAST 1a monomer is indicated in dark blue and TCAST 1b in light blue. Direct duplicated regions which are also the sites of site-specific recombination are indicated by yellow lines. B) Based on proposed model, possible formation of element 2 presented as a monomer form a dimer could be, as well, result of intramolecular homologous recombination

Analysis of gene expression performed in this study did not show any significant influence of satellite-like elements on gene regulation in normal condition. However, we have observed the difference in expression of genes associated with satellite elements under stress conditions. Transcriptional responses to different environmental challenges have been documented for many euchromatin-associated genes and/or gene networks related to different organisms (Rizzi et al., 2004; Pecinka et al., 2010). These reports have been associated with alternations in epigenetic regulatory mechanisms, such as changes in the distribution of DNA methylation, histone modifications or populations of regulatory small RNAs. The involvement of small RNAs leading to modifications of epigenetics marks at target genes and/or degradation of mRNA or translational inhibition by post-transcriptional gene silencing seem to play important roles in stress responses (Tittel-Elmer et al., 2010). SatDNA-derived siRNA have been implicated in promoting heterochromatin formation through RNA interference (RNAi) mechanism by guiding chromatin modifiers such as histone methylase that induces H3K9 methylation (Verdel et al., 2004). The RNAi machinery has been shown to be evolutionary conserved and connected with regulation at the chromatin level in *S. pombe* (Volpe et al., 2002; Verdel and Moazed, 2005), *Tetrahymena* (Berstein and Allis, 2005), *C. elegans* (Aravin and Tuschl, 2005), *Arabidopsis* (Lippman and Martienssen, 2004), *Drosophila* (Pal-Bhadra et al., 2004, Vagin et al., 2006) and vertebrates (Lippman and Martienssen, 2004). In addition to aspect of transcriptional silencing through chromatin modifications, RNAi mechanism was implicated in posttranscriptional silencing of retrotransposones and tandem repeats (Lippman and Martienssen, 2004). In *Drosophila*, in addition of piRNA which carry the function of silencing repetitive elements predominantly in germ line, new class of small RNAs has been recently discovered (Czech et al., 2008; Okamura and Lai, 2008) named endogenous small interfering RNAs (endo-siRNAs). Endo-siRNAs derive from transposons, heterochromatic sequences, intergenic regions, long RNA transcripts with extensive structure and from mRNA. It is proposed that there is a cross-talk

between the piRNAs and endo-siRNAs in silencing repetitive elements (Ghildiyal and Zamore, 2009; Bühler et al., 2007).

In *Tribolium*, after heat stress, strong increase of TCAST related siRNAs has been detected (Pezer and Ugarković, 2012). Peak of siRNAs was correlated with increase in H3K9 methylation of TCAST satellite element located in centromeric and pericenromeric regions. In addition to participating in heterochromatin formation in centromeric and pericenromeric regions, it is possible that siRNAs originating from satellite repeats have a more extensive role in expression of genes that are embedded with TCAST satellite-like element. After heat stress we have observed downregulation of both genes tested disregarding whether they had TCAST element or not. Observed downregulation could be a consequence of general impact of heat shock on nuclear processes (Richter et al., 2010). However, difference in expression of gene observed after one hour of recovery seems to be related to presence of TCAST element. Only genes that contained the TCAST insert remained downregulated.

Influence of heterochromatin on expression of gene located in is vicinity is very well documented in *Drosophila* (Talbert and Henikoff, 2006). It was established that during embryogenesis heterochromatin acquires ability to silent genes approximately an hour after heterochromatin is first visible in cleavage stage embryos (Lu et al., 1998). It is possible that similar mechanism acts in *Tribolium* after heat stress and that temporary increase of siRNAs related to TCAST satellite elements induces silencing of genes that contain TCAST insert through chromatin modifications. TCAST element, inserted in the intron, could in that case serve as a seeding region for establishing and spreading of repressive chromatin marks through recruitment of siRNA. TCAST related siRNA, which are upregulated after heat shock, could function in a sequence-specific targeting of TCAST element situated in the intron of gene and guiding chromatin modifiers, primarily histone methylase in heterochromatin assembly (Figure 17).

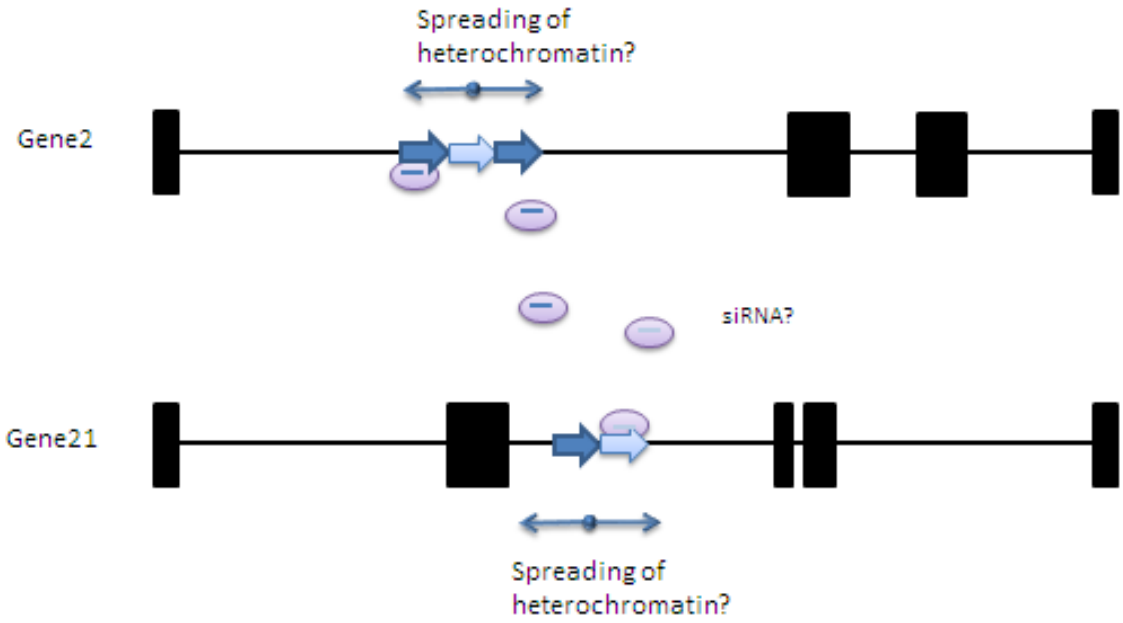


Figure 17. Model for TCAST siRNA-dependent silencing of genes that contain TCAST element. TCAST related siRNA, could function in a sequence-specific targeting of TCAST element situated in the intron of gene and in guiding chromatin modifiers primarily histone methylase in heterochromatin assembly. TCAST elements are represented by arrows and siRNA are in corresponding color as their complementary TCAST element.

5. Conclusion

This study reveals polymorphism of dispersed TCAST satellite-like elements among strains of *T. castaneum* and gives a strong indication of the site specific recombination role as a mechanism of dispersion of repetitive elements.

The study also indicates the influence of dispersed TCAST satellite-like elements on the regulation of genes located in their vicinity under specific heat stress conditions. Further studies are necessary to confirm this phenomenon and to explain the possible role of dispersed repetitive elements in the evolution of gene regulation.

6. References

Aravin A., Tuschl T. (2005): Identification and characterization of small RNAs involved in RNA silencing. FEBS Lettters 579: 5830-5840

Berstein E., Allis D. C. (2005): RNA meets chromatin. Genes Dev. 19: 1635-1655

Brajković J., Feliciello I., Bruvo-Mađarić B., Ugarković D. (2012): Satellite DNA-like elements associated with genes within euchromatin of the beetle Tribolium castaneum. G3 2: 931-41.

Brown S. J., Denell R. E., Beeman R. W. (2003): Beetling around the genome. Genet. Res. 82, 155–161

Brown J. D., Mitchell S. E., O'Neill R. J. (2012): Making a long story short: noncoding RNAs and chromosome change. Heredity 18: 42-49

Bustin S. A., Benes V., Garson J. A., Hellemans J., Huggett J., Kubista M., Mueller R., Nolan T., Pfaffl M. W., Shipley G. L., Vandesompele J., Wittwer C. T. (2009): The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-time PCR Experiments. Clinical Chemistry 55: 611-622

Bühler M., Haas W., Gygi S. P., Moazed D. (2007): RNAi-dependent and -independent RNA turnover mechanisms contribute to heterochromatic gene silencing. Cell 129: 707-21.

Cohen S., Regev A., Lavi S. (1997): Small polydispersed circular DNA (spcDNA) in human cells: association with genomic instability. Oncogene, 14: 977-985

Cohen S., Menut S., Méchali M. (1999): Regulated formation of extrachromosomal circular DNA molecules during development in *Xenopus laevis*. Mol Cell Biol., 19: 6682-6689.

Czech B., Malone C. D., Zhou R., Stark A., Schlingeheyde C., Dus M., Perrimon N., Kellis M., Wohlschlegel J. A., Sachidanandam R., Hannon G. J. , Brennecke J. (2008): An endogenous small ineterfering RNA pathway in Drosophila. Nature 453: 798-802

Davidson E. H., Britten R. J. (1979): Regulation of gene expression: possible role of repetitive sequences. Science 204: 1052-1059

Feliciello I., Picariello O., Chinali G. (2005): The first characterisation of the overall variability of repetitive units in a species reveals unexpected features of satellite DNA. Gene 349: 153-64.

Feliciello I., Picariello O., Chinali G. (2006): Intra-specific variability and unusual organization of the repetitive units in a satellite DNA from Rana dalmatina: molecular evidence of a new mechanism of DNA repair acting on satellite DNA. Gene 383: 81-92.

Feliciello I., Chinali G., Ugarković Đ. (2011): Structure and population dynamics of the major satellite DNA in the red flour beetle Tribolium castaneum. Genetica 139: 999-1008

Ghildiyal M., Zamore P. D. (2009): Small silencing RNAs: an expanding universe. Nature Reviews Genetics 10: 94-107

Howe R. W. (1956): The effect of temperature and humidity on the rate of development and mortality of *Tribolium castaneum* (Herbst) (coleoptera, tenebrionidae). Annals of Applied Biology 44: 356–368

Karlen Y., McNair A., Perseguers S., Mazza C., Mermod N. (2007): Statistical significance of quantitative PCR. BMC Bioinformatics 8: 131.

Lippman Z., Martienssen R. (2004): The role of RNA interference in heterochromatic silencing. Nature 431: 364-370

Lord J. C., Hartzer K., Toutges M., Oppert B. (2009): Evaluation of quantitative PCR reference genes for gene expression studies in *Tribolium castaneum* after fungal challenge. Journal of Microbiological Methods 80: 219-221

Lorenzen M. D., Doyungan Z., Savard J., Snow K., Crumly L. R., Shippy T. D., Stuart J. J., Brown S. J., Beeman R. W. (2005): Genetic Linkage Maps of the Red Flour Beetle, *Tribolium castaneum*, Based on Bacterial Artificial Chromosomes and Expressed Sequence Tags. Genetics 170: 741–747

Lu Y. B., Ma J., Eissenberg J. C. (1998): Developmental regulation of heterochromatin-mediated gene silencing in Drosophila. Development 125: 2223-2234

Meister G., Tuschl T. (2004): Mechanism of gene silencing by double-stranded RNA. Nature 431: 343-349.

Okamura K., Lai E. C. (2008): Endogenous small interfering RNAs in animals. Nat Rev Mol Cell Biol 9: 673-678.

Pal-Bhadra M., Leibovitch B. A., Gandhi S. G., Rao M., Bhadra U., Birchler J. A., Elgin S. C. R. (2004): Heterochromatic Silencing and HP1 Localization in Drosophila Are Dependent on the RNAi Machinery. Science 303: 669-671

Palomeque T., Lorite P. (2008): Satellite DNA in insects: a review. Heredity 100: 564-573

Pecinka A., Dinh H. Q.,Baubec T., Rosa M., Lettner N., Scheid M. O. (2010): Epigenetic Regulation of Repetitive Elements is Attenuated by Prolonged Heat Stress in *Arabidopsis*. The Plant Cell 22: 3118-3129

Pezer Ž., Ugarković Đ. (2012): Satellite DNA-associated siRNAs as mediators of heat shock response in insects. RNA Biology 9: 587-595.

Pezer Ž., Brajković J., Feliciello I., Ugarković Đ. (2012): Satellite DNA-Mediated Effects on Genome Regulation: Genome Dyn. 7: 153-169

Pont G., Degroote F., Picard G. (1987): Some extrachromosomal circular DNAs from *Drosophila* embryos are homologous to tandemly repeated genes. J Mol Biol. 195: 447-451.

Ramakers C., Ruijter J. M., Deprez RH, Moorman A. F. (2003): Assumption-free analysis of quantitative real-time PCR data. Neurosci Letters 339: 62-66

Richteer K., Haelbeck M., Buchner J. (2010): The Heat Shock Response: Life on the Verge of death. Molecular Cell 40 :253-264

Rizzi N., Denegri M., Chiodi I., Corioni M., Valgardsdottir R., Cobianchi F., Riva S., Biamonti G. (2004): Transcriptional activation of a constitutive heterochromatic domain of the human genome inresponse to heat shock. Mol Biol Cell. 15: 543-51

Ruijter J. M., Ramakers C., Hoogaars W. M., Karlen Y., Bakker O., van den Hoff M. J., Moorman A. F. (2009): Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. Nucleic Acids Research 37: e45

Ruijter J. M., Pfaffl M. W., Zhao S., Spiess N. A. Boggy G., Blom J., Rutledge R. G., Sisti D., lievens A., De Preter K.,Derveaux S.,Hellemans J., Vandesompele J. (2012): Evaluation of qPCR curve analysis methods for reliable biomarker discovery: Bias, resolution, precision, and implications. Methods xxx: xxx-xxx

Talbert P. B., Henikoff S. (2006): Spreading of silent chromatin: inaction at a distance. Nat Rev Genet 7: 793–803

Tittel-Elmer M., Buscher E., Broger L., Mathieu O., Paszkowski J., Vailiant I. (2010): Stress-Induced Activation of Heterochromatic Transcription. PLoS Genetics 6: e1001175

Tomoyasu Y., Miller S. C., Tomita S., Schoppmeier M., Grossman D., Buchner G. (2008): Exploring systemic RNA interference in insects; a genome-wide survey for RNAi in *Tribolium*. Genome Biology 9: R10

Toutges M. J., Hartzer K., Lord J., Oppert B.(2010): Evaluation of Reference Genes for Quantitative Polymerase Chain Reaction across Life Cycle Stages and Tissue Types of *Tribolium castaneum*. J. Agric. Food Chem. 58: 8948-8951

Tribolium Genome Sequencing Consortium (2008): The genome of the model beetle and pest *Tribolium castaneum*. Nature 452: 949-955.

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

Ugarković Đ. (2005): Functional elements residing within satellite DNAs. EMBO Rep 6: 1035-1039Vagin V. V., Sigova A., Chengjian L., Seitz H., Gvozdev V., Zamore P. D. (2006) A Distinct Small RNA Pathway Silences Selfish Genetic Elements in the Genome. Science 313: 320-324

Verdel A., Jia S., Gerber S., Sugiyama T., Gygi S., Grewal S. I. S., Moazed D. (2004): RNAi-Mediated Targeting of Heterochromatin by the RITS Complex. Nature 303: 672-676

Verdel A., Moazed D. (2005): RNAi-directed assembly of heterochromatin in fission yeast. FEBS Letters 579: 5872-5878

Volpe T. A., Kidner C., Hall I. M., Teng G., Grewel S. I. S., Martienssen R. A. (2002): Regulation of Heterochromatic Silencing and Histone H3 Lysine-9 Methylation by RNAi. Science 297: 1833-1837

Wong M., Medrano J.F. (2005): Real-time PCR for mRNA quantification. BioTechniques 39: 75-85

Wang S., Lorenzen M. D., Beeman R. W., Brown S. J.(2008): Analysis of repetitive DNA distribution patterns in the *Tribolium castaneum* genome. Biology 9: R61

Yuan J. S., Reed A., Chen F., Steward N. J. (2006): Statistical analysis of real time PCR data. BMC Bioinformatics 7: 85is