

Research Note

Distribution of DTHS3 satellite DNA across 12 bivalve species

Eva Šatović^{1*} and Miroslav Plohl¹

¹ Division of Molecular Biology, Ruđer Bošković Institute, Bijenička 54, 10 000 Zagreb, Croatia

* Corresponding Author:

Eva Šatović, Division of Molecular Biology, Ruđer Bošković Institute, Zagreb, Croatia, e-mail: esatovic@irb.hr

Running (short) title: DTHS3 satellite DNA

Keywords: satellite DNA, Bivalvia, evolution

Abstract

In this work, characterization of DTHS3 satellite DNA was further expanded within the Class Bivalvia. Monomer variants of DTHS3 satDNA were compared in 12 bivalve species belonging to two different Subclasses, Heterodonta and Pteriomorpha. This satDNA, whose age is estimated to a minimum of 516 Ma, is contributing to the concept of the dual character of satDNA sequences: their sequence preservation throughout long evolutionary periods and generation of species-specific variants of the same satDNA family.

Introduction

Class Bivalvia is represented by a large number of marine and freshwater mollusks, found in aquatic habitats throughout the world. Despite their significant ecological and commercial value, genomic studies on these organisms are not abundant, and information is still limited to a few species. Genome sequencing projects have been performed only for the pearl oyster *Pinctada fucata* (Takeuchi et al. 2012), pacific oyster *Crassostrea gigas* (Zhang et al. 2012), and filter-feeder mussel *Mytilus galloprovincialis* (Murgarella et al. 2016).

Significant fraction of all eukaryotic genomes is made of repetitive sequences. Among them are satellite DNA (satDNA) sequences, organized in arrays of tandemly repeated monomers (Miklos 1985). *Donax trunculus* is an example of a bivalve species harboring a number of different satDNAs: DTE (Plohl and Cornudella 1996), DTHS1, DTHS2, DTHS3, DTHS4 (Plohl and Cornudella 1997), DTF1 (Petrović and Plohl 2005), DTF2 (Petrović et al. 2009) and BIV160 (Plohl et al. 2010). At the same time, some satDNAs can be shared among closely (*phBglIII400*, Passamonti et al. 1998), or even distantly related species (BIV160, Plohl et al. 2010). DTHS3 satDNA was originally discovered in *D. trunculus* after HindIII digestion of genomic DNA (Plohl and Cornudella 1997). Experimental procedure based on restriction enzyme digestion generated a regular ladder of genomic fragments, a feature typical for sequences repeated in tandem. DTHS3 was therefore characterized as a satDNA. It has 145 bp-long monomer sequence which holds several copies of GGTC pentanucleotide motif and comprises 0,035% of *D. trunculus* genome. Presence of this satellite DNA in species *Ruditapes decussatus* and *R. philippinarum* has been recently reported in Šatović et al. (2016).

Here, we report the presence of DTHS3 satDNA in 9 additional bivalve species, further expand its characterization, and estimate its age to a minimum of 516 Ma.

Materials and methods

Genomic DNAs of *D. trunculus*, *R. decussatus* and *R. philippinarum* were partially digested and used for preparation of genomic libraries, followed by generation of small local databases enriched in repetitive DNA sequences (details in Šatović and Plohl 2013 and Šatović et al. 2016). This methodology followed the approach of Biscotti et al. (2007) based on the fact that during hybridization process, nucleotide sequences that are present in the genome in multiple copy number will give more intense hybridization signals compared to the single-copy ones. Repetitive sequences belonging to DTHS3 satDNA have previously been deposited in GenBank in course of more detailed studies on repetitive sequences in these organisms (Šatović and Plohl 2013, Šatović et al. 2016). Here, GenBank search showed its presence in additional bivalve species. Sequences used during this study have the following accession numbers: KU682294, KU682297, KU682299, AM852203.1, AM871347.1, AM851753.1, KU682284, KU682285, X94611.1, X94540.1, X94542.1, KU682312, KU682313, KU682310, AM872828.1, AM874475.1, AF108933.1, FY999295.1, FY999259.1, FY999264.1, GR210966.1, GR211262.1, HO206039.1, AB231865.1, EU054059.1, GO309863.1, GO308956.1,

GO309030.1, GO308318.1, GO310525.1, AM853355.1, AM853502.1, EY436206.1, EY436734.1, EY435449.1, EY436295.1, EY436878.1, EY434278.1, EY436850.1, EY433829.1, EY435144.1, EY435174.1, EY433716.1, EY435319.1, MWPT01000253.1 and 413002.

DTHS3 monomer frame was determined in relation to previously reported monomers in Plohl and Cornudella (1997). For phylogenetic analysis, monomers have been aligned using the Muscle algorithm implemented in Mega v. 5.05 (Tamura et al. 2011). The same program was used to determine the best substitution models and to generate Maximum Likelihood trees; nodal support has been obtained with the bootstrap of 500 replicates. Phylogenetic tree showing the relationships between species was build based on sequences derived from cytochrome oxidase subunit 1 (COI) genes, holding accession numbers: AB076943.1, AB076949.1, EU484436.1, FJ434678.1, FJ434679.1, JX051541.1, KC429143.1, KP099052.1, KU252878.1, KU905884.1, KU905937.1, KU906094.1. Rest of the sequence alignments and sequence analysis was performed with Geneious 9.1.7 program (Biomatters Ltd.).

Results

For the purpose of more detailed investigation of DTHS3 satDNA, this sequence was extracted from local databases obtained earlier for three bivalve genomes (Šatović and Plohl 2013) and complemented with results obtained after a screen of GenBank against this sequence. This search yielded 93 monomers across 10 bivalve species: *D. trunculus*, *R. decussatus*, *R. philippinarum*, *Sinonovacula constricta*, *Mercenaria mercenaria*, *Meretrix lusoria*, *M. meretrix*, *Spisula solidissima*, *Crassostrea gigas*, *C. virginica*, *Mytilus galloprovincialis* and *Dreissena bugensis*; their classification can be seen in Table 1. In the analyzed set of sequences, longest array of DTHS3 was found in *S. solidissima*, containing 6 and 8 complete monomers. In other species up to 4 complete monomers were found repeated in tandem, not excluding the possibility that longer arrays exist in the genome.

First report on DTHS3 satellite in *D. trunculus* was brought by Plohl and Cornudella (1996), yielding 3 complete and 1 truncated monomer sequence. Our local database search determined its presence in 2 more genomic clones, 84_19 and DTC32Alu, the latter harboring an array of 4 monomers repeated in tandem. Complete monomers obtained from this species exhibit nucleotide sequence similarity between 83.2 - 98.7%.

For *R. decussatus*, local database search found presence of this sequence in 5 genomic clones. In the case of D12 (KU682294), monomer was interrupted by insertion of SINE element RUDI (Luchetti et al. 2016), leaving the rest of the monomer sequence intact. In addition, we found 3 sequences in EST database, and altogether 8 complete monomers were extracted. Similarity between monomers from this species varies between 61.5 and 96.6%.

In *R. philippinarum* DTHS3 was found in 2 genomic clones and 2 more sequences that belong to the EST database, yielding altogether 10 complete monomers. Overall monomer similarity varies between 61% and 95.2%.

EST database search in the case of *S. constricta* yielded 6 sequences from which 11 complete monomers were extracted, showing similarity between 55.1 and 85.5%, while for *M. meretrix* 4 EST database hits have yielded 7 DTHS3 monomers, with sequence similarity between 60.0 and 76.5%.

In *M. mercenaria* only one fragment holding one truncated and two complete DTHS3 monomers was found, latter ones sharing 83.5% similarity. The repeats are here followed by a microsatellite sequence, similarly as in DTC84 (Šatović and Plohl 2013) and *pearl* mobile elements (Gaffney et al. 2003).

Four *M. lusoria* cDNA fragments harbor DTHS3 sequence, from which 9 complete highly similar monomers were extracted (86.9 - 100%). As three out of four cDNA sequences show high overall sequence similarity, it is possible that they are transcripts from the same genomic location.

GenBank database search yielded two genomic fragments of *S. solidissima* containing arrays of this satellite DNA. Similarity among complete monomers is 69.2 - 97.9%. In the case of AB231865.1 genomic fragment, DTHS3 array is again followed by a microsatellite sequence.

For *D. bugensis*, DTHS3 sequence was found within 12 fragments, giving 15 monomers whose nucleotide similarity varies between 57.5 and 100%.

Four *C. gigas* cDNA fragments contained DTHS3 sequence, from where four monomers were extracted, exhibiting 55.1 - 78.9% of similarity among each other.

Interesting organization of DTHS3 repeats is found in *C. virginica* and *M. galloprovincialis*. One scaffold from *C. virginica* WGS base holds two DTHS3 repeats followed by a microsatellite, while *M. galloprovincialis* WGS contig harbors three monomers in tandem, also followed by a microsatellite. Genomic fragments from the two species share sequence similarity

among monomers and the microsatellite, as well as in 505 bp preceding the repeats. Altogether, they exhibit 96,5% identity over 836 bp, not taking into consideration the central repeat of the array belonging to *M. galloprovincialis*, which holds a 19 bp insertion. In addition, behind 836-bp fragment, *C. virginica* harbors another monomer and a half, positioned 335 bp downstream of the end of the microsatellite segment. Within each species, monomers derived from *C. virginica* exhibit 73.8 - 86.9% similarity, while those derived from *M. galloprovincialis* have 78.0 - 86.2% similarity. Blast search against *C. gigas* genome, closely related to *C. virginica*, did not show presence of 836 bp sequence segment present in *C. virginica* and *M. galloprovincialis*.

As mentioned, in *M. mercenaria* and *S. solidissima* DTHS3 monomers are also found followed by a microsatellite, but only in case of *C. virginica* and *M. galloprovincialis* strong homology in sequences preceding and following the repeats are observed.

Phylogenetic analysis revealed 2 patterns. One is species-specific clustering of monomers, as noticeable for *S. solidissima*, *M. mercenaria*, *D. trunculus* and *S. constricta*. Second pattern groups monomers derived from different species, as observed for *R. decussatus* and *C. gigas*, *C. virginica* and *M. galloprovincialis* and for *R. philippinarum*, *M. meretrix* and *M. lusoria* (Fig. 1a). *D. bugensis* monomers partially group together, and at the same time cluster with those from *C. virginica* and *M. galloprovincialis*. Monomers derived from closely related species can be found intermingled together, as for *M. lusoria* and *M. meretrix*. In the same time, monomers derived from *R. decussatus* and *R. philippinarum* group closer to those belonging to more distantly related species than to each other. In that respect, monomers derived from *R. decussatus* and *C. gigas* cluster together, although these species belong to two different Subclasses (Table 1, Figure 1b). The same accounts for monomers of *C. virginica* and *M. galloprovincialis*, belonging to different orders, with exception of the monomer positioned outside of the 836 bp long fragment of *C. virginica*, which positions closer to monomers from *D. bugensis*.

Discussion

Satellite repeats belonging to DTHS3 family have been detected in 12 species belonging to the two subclasses of the Class Bivalvia, Heterodonta and Pteriomorpha. As estimated time of diversification for Pteriomorpha is 516 Ma (Bieler et al. 2014), this satDNA can be dated very closely to BIV160 satDNA, estimated to appear 540 Ma and representing the oldest described

satDNA so far (Plohl et al. 2010). One of the potential ways of genomic dispersal of BIV160 sequences was proposed to be linked to Miniature inverted-repeat transposable elements (MITEs), based on sequence similarity with the *pearl* family of elements found in oysters (Gaffney et al., 2003). These elements are composed of a central region holding different number of monomer-like repeats followed by a microsatellite segment, while the whole structure is flanked by sequences holding short inverted repeats. The same structural organization has been found in DTC84 element of *D. trunculus* (Šatović and Plohl 2013). Similarly structured tandem repeat-carrying elements also exist in evolutionary distant organisms such as *Drosophila*, and are altogether categorized as non-autonomous eukaryotic rolling-circle transposable elements of the Helitron / Helentron superfamily (Thomas and Pritham, 2014).

In this study, DTHS3 repeats in *M. mercenaria*, *S. solidissima*, *C. virginica* and *M. galloprovincialis* have been found in arrays that are followed by a microsatellite sequence, opening the possibility that these repeats also can be found within structures similar to that of Helitron / Helentron mobile elements (Gaffney et al. 2003, Šatović and Plohl 2013).

Obtained results suggest that DTHS3 behaves quite similar to BIV160 satDNA (Plohl et al. 2010) in estimated age, sequence conservation, and potential to generate novel species-specific subfamilies. In both satDNAs, observed results indicate presence of species-specific subfamilies, together with interspecifically intermingled monomer variants (Fig. 1). Most probably, diversification of species-specific variants from a common ancestral sequence has occurred following or in course of speciation (Plohl et al. 2010).

Existence of the 836 bp-long extremely similar fragments in *C. virginica* and *M. galloprovincialis* is intriguing, as the same sequence was not found in *C. gigas*, closely related to *C. virginica*. There are three possible explanations for this observation: the sequence existed in the common Pteriomorphia ancestor but in course of time was lost in the genome of *C. gigas*; the sequence is present in *C. gigas* genome but it was omitted in the genome sequencing process; a horizontal transfer event occurred. In addition, the discrepancy in sequence grouping and known mollusk systematics (Table 1, Figure 1b) might indicate a horizontal transfer as a possible way by which DTHS3 monomers became shared between *R. decussatus* and *C. gigas*, distant species that belong to different subclasses. Existing overlap in their geographical distribution along British Islands, Iberian Peninsula and the Mediterranean (information retrieved from <http://www.fao.org/home/en/>) and ancient divergent time between two species (516 Ma, Bieler et al. 2014) would go in favor of this hypothesis.

In conclusion, DTHS3 is a satDNA family that is widely spread among bivalve species. It persisted throughout long evolutionary periods, and is probably one representative of ancient repetitive sequences that were present in a common ancestor of Bivalve species. Thus, diversification of DTHS3, together with different potential ways of its propagation, have influenced and contributed to the structuring of bivalve genomes on a large evolutionary scale.

References

- Bieler R., Mikkelsen P. M., Collins T., et al. 2014 Investigating the Bivalve Tree of Life – an exemplar-based approach combining molecular and novel morphological characters. *Invertebr. Syst.* 28, 32–115.
- Biscotti M. A., Canapa A., Olmo E., Barucca M., Teo C. H., Schwarzacher T., Dennerlein S., Richter R., Heslop-Harrison J. S. P. 2007 Repetitive DNA, molecular cytogenetics and genome organization in the King scallop (*Pecten maximus*). *Gene* 406, 91–98.
- Gaffney P. M., Pierce J. C., Mackinley A. G., Titchen D. A., Glenn W. K. 2003 *Pearl*, a novel family of putative transposable elements in bivalve mollusks. *J. Mol. Evol.* 56, 308–316.
- Luchetti A., Šatović E., Mantovani B., Plohl M. 2016 RUDI, a short interspersed element of the V-SINE superfamily widespread in molluscan genomes. *Mol. Genet. Genomics* 291, 1419–1429.
- Meštrović N., Castagnone-Sereno P., Plohl M. 2006 Interplay of selective pressure and stochastic events directs evolution of the MEL172 satellite DNA library in root-knot nematodes. *Mol. Biol. Evol.* 23, 2316–2325.
- Miklos G. L. G. 1985. Localized highly repetitive DNA sequences in vertebrate and invertebrate genomes. In *Molecular Evolutionary Genetics*. (ed: R. J. MacIntyre), pp 241–321. Plenum Press, New York.
- Murgarella M., Puiu D., Novoa B., Figueras A., Posada D., Canchaya C. 2016 A First Insight into the Genome of the Filter- Feeder Mussel *Mytilus galloprovincialis*. *PLoS One* 11, 1–22.
- Passamonti M., Mantovani B., Scali V. 1998 Characterization of a highly repeated DNA family in tapetinae species (mollusca bivalvia: veneridae). *Zoolog. Sci.* 15, 599–605.
- Petrović V., Pérez-García C., Pasantes J. J., Šatović E., Prats E., Plohl M. 2009 A GC-rich satellite DNA and karyology of the bivalve mollusk *Donax trunculus*: a dominance of GC-rich heterochromatin. *Cytogenet. Genome Res.* 124, 63–71.
- Petrović V., Plohl M. 2005 Sequence divergence and conservation in organizationally distinct subfamilies of *Donax trunculus* satellite DNA. *Gene* 362, 37–43.

- Plohl M., Cornudella L. 1996 Characterization of a complex satellite DNA in the mollusc *Donax trunculus*: analysis of sequence variations and divergence. *Gene* 169, 157–164.
- Plohl M., Cornudella L. 1997 Characterization of interrelated sequence motifs in four satellite DNAs and their distribution in the genome of the mollusc *Donax trunculus*. *J. Mol. Evol.* 44, 189–198.
- Plohl M., Petrović V., Luchetti A., Ricci A., Šatović E., Passamonti M., Mantovani B. 2010 Long-term conservation vs high sequence divergence: the case of an extraordinarily old satellite DNA in bivalve mollusks. *Heredity*. 104, 543–551.
- Salser W., Bowen S., Browne D., el-Adli F., Fedoroff N., Fry K., Heindell H., Paddock G., Poon R., Wallace B., Whitcome P. 1976 Investigation of the organization of mammalian chromosomes at the DNA sequence level. *Fed. Proc.* 35, 23–35.
- Šatović E., Plohl M. 2013 Tandem repeat-containing MITE elements in the clam *Donax trunculus*. *Genome Biol. Evol.* 5, 2549–2559.
- Šatović E., Vojvoda Zeljko T., Luchetti A., Mantovani B., Plohl M. 2016 Adjacent sequences disclose potential for intra-genomic dispersal of satellite DNA repeats and suggest a complex network with transposable elements. *BMC Genomics* 17, 997.
- Takeuchi T., Kawashima T., Koyanagi R., Gyoja F., Tanaka M., Ikuta T., Shoguchi E., Fujiwara M., Shinzato C., Hisata K., et al. 2012. Draft genome of the pearl oyster *Pinctada fucata*: a platform for understanding bivalve biology. *DNA Res.* 19, 117–130.
- Tamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S. 2011 MEGA5 : Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molec. Biol. Evol.* 28, 2731–2739.
- Thomas J., Pritham E. J. 2015 Helitrons, the Eukaryotic Rolling-circle Transposable Elements. *Microbiol. Spectr.* 3, 1–32.
- Zhang G., Fang X., Guo X., Li L., Luo R., Xu .F, Yang P., Zhang L., Wang X., Qi H., et al. 2012 The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 490, 49–54.

Table 1: Classification of 12 bivalve species used in this study by World Register of Marine Species (<http://www.marinespecies.org/index.php>)

Kingdom	Phylum	Class	Subclass	Order	Superfamily	Family	Genus	Species	
Animalia	Mollusca	Bivalvia	Heterodonta	Cardiida	Tellinoidea	Donacidae	Donax	<i>Donax trunculus</i>	
				Venerida	Veneroidea	Veneridae	Ruditapes	<i>Ruditapes decussatus</i>	
								<i>Ruditapes philippinarum</i>	
							Mercenaria	<i>Mercenaria mercenaria</i>	
							Meretrix	<i>Meretrix lusoria</i>	
								<i>Meretrix meretrix</i>	
					Mactroidea	Mactridae	Spisula	<i>Spisula solidissima</i>	
				Adapedonta	Solenoidea	Pharidae	Sinonovacula	<i>Sinonovacula constricta</i>	
				Myida	Dreissenoidea	Dreissenoidae	Dreissena	<i>Dreissena bugensis</i>	
				Pteriomorphia	Ostreida	Ostreoidea	Ostreidae	Crassostrea	<i>Crassostrea gigas</i>
									<i>Crassostrea virginica</i>
					Mytilida	Mytiloidea	Mytilidae	Mytilus	<i>Mytilus galloprovincialis</i>

Figure description

Figure 1:

a) Maximum likelihood bootstrap consensus tree built on 93 DTHS3 monomers from 12 bivalve species. b) Maximum likelihood phylogenetic tree based on COI sequences showing relationships between species.

