Characteristics and evolution of satellite DNA sequences in bivalve mollusks

E. ŠATOVIĆ, T. VOJVODA ZELJKO, & M. PLOHL

Division of Molecular Biology, Ruđer Bošković Institute, Zagreb, Croatia

(Received 20 December 2017; accepted 13 February 2018)

Abstract

Mollusks of the class Bivalvia have attracted attention because of the extraordinarily important roles they play in marine ecosystems and in aquaculture. Data obtained from genetic studies performed on these species are accumulating rapidly, particularly in recent years when several genomic and transcriptomic studies have been carried out, or are in progress. Despite this, knowledge concerning satellite DNAs, tandemly repeated non-coding genomic sequences important for comprehending genomic architecture and function as a whole, is fragmentary and limited to a relatively small number of mollusk species. Here, we present an overview of the studied satellite DNAs and their characteristics in bivalve mollusks, and discuss the implications of these results. In addition to the general features common for these sequences, bivalve satellite DNAs show some distinct specificities which may be intriguing for the broad scientific community involved in unravelling repetitive genome components. The most striking are low genomic contribution, diversity of sequence families, extremely old ancestry, links with mobile elements, and unusual methylation patterns. Although current results were obtained in classical studies on individual species and their satellite DNA families, it can be postulated that they defined fundamental characteristics of these sequences in bivalve species generally, and will be further explored in detail by future satellitome and other high-throughput studies.

Keywords: Bivalves, mollusks, satellite DNA, repetitive DNA sequences, evolution

Introduction

The class Bivalvia consists of 9200 species found in marine and freshwater habitats throughout the world. They play an important role in the ecosystems they inhabit and serve as support to many mollusk and other communities, participating in the transfer of minerals and organic matter to benthic habitats. As a result of filtration during the feeding process, they can accumulate various xenobiotic substances, which make them also organisms of choice for environmental monitoring (Gosling 2003). In accordance with their extreme ecological and commercial importance in marine ecosystems and aquaculture, interest in genetic research on these organisms is growing steadily, and has increased particularly in recent years. Linkage maps, transcriptomics and proteomics studies have been carried out to reveal the genetic and molecular basis of traits of interest to the bivalve farming industry, mainly disease susceptibility, tolerance to environmental stress, and processes connected to metabolism and growth (Saavedra & Bachère 2006). Previous research on these organisms was largely the result of interest in specific gene functions, while recent studies have been moving towards genome-wide analyses. Genome sequencing projects have so far been performed only for the pearl oyster Pinctada fucata (Gould, 1850) (Takeuchi et al. 2012), the pacific oyster Crassostrea gigas (Thunberg, 1793) (Zhang et al. 2012) and the filter-feeder mussel Mytilus galloprovincialis Lamarck, 1819 (Murgarella et al. 2016), and sequenced and assembled genomes of other bivalve species can be anticipated in the near future.

Eukaryotic genomes contain a significant portion of non-coding DNA represented by a large amount of different classes of repetitive sequences, found either interspersed or organised in tandem (López-Flores & Garrido-Ramos 2012; Biscotti et al. 2015). Satellite DNAs (SatDNAs) are the most abundant
Satellite DNA in bivalves

Satellite DNA sequences are usually AT rich, and their abundance in genomes shows a wide range of variation, with values from less than 0.5% to more than 50%. Due to the phenomenon of concerted evolution, sequence variability among monomers within a satDNA family is low, usually 2–3% (reviewed by Plohl et al. 2012; Garrido-Ramos 2017). Although many unrelated satellites can populate genomes and genomic satDNA profiles evolve rapidly, their DNA sequences can remain conserved over long evolutionary periods (reviewed by Plohl et al. 2008, 2012). Among diverse families, satDNA monomer length can vary significantly, but 150 to 210-bp-long monomers predominate. They are considered to be evolutionarily favoured (Heslop-Harrison & Schwarzacher 2013), as they correspond to the nucleosomal length (Henikoff et al. 2001). SatDNA monomers sometimes contain about 30 bp long sequence segments that exhibit reduced variability with respect to the rest of the monomer sequence. Such conserved segments can facilitate recombination between satDNA monomers and sequence homogenisation (Kuhn et al. 2009) or have some other functional roles, for example, in DNA–protein interactions. One example is the CENP-B protein, which facilitates centromere formation and plays an important role in the assembly of centromere-specific structures upon recognising and binding the CENP-B box within the higher primate alpha-satellite sequence (Masumoto et al. 1989). Satellite DNAs have also shown strong relationships with mobile elements, in addition to being located in close proximity to each other in certain genomic regions. For example, tandem repeats are detected as structural components of some mobile elements or satellite repeats can be derived from parts of mobile elements (reviewed by Mestrovic et al. 2015).

The commonly used method employed in the detection of satellite sequences is digestion of genomic DNA with restriction endonucleases, followed by electrophoretic separation. A characteristic ladder pattern is the result of variability in the restriction site within satellite monomers organised in an array. Alternatively, construction of partial genomic libraries followed by colony-lift hybridisation with fragmented total genomic DNA is sometimes used as a strategy in detecting repetitive genomic fragments, including satDNAs (e.g. Biscotti et al. 2007).

Difficulties in sequencing and assembly of arrays composed of highly similar tandem repeats have resulted in the fact that satDNA sequences are seriously underrepresented in genome sequencing outputs. With respect to bivalve species, genomic projects are no exception, and have not yielded comprehensive information regarding satDNAs. Predictions are that 202 Mb (36% of the genome) of Crassostrea gigas genome is enriched in repetitive sequences (Zhang et al. 2012), but over 62% of the detected repeats could not be assigned to known categories. In the case of the oyster Pinctada fucata (Takeuchi et al. 2012), about 7.9% of the genome is occupied by tandem repetitive elements, among which 3.7% represents microsatellite sequences. In Mytilus galloprovincialis (Murgarella et al. 2016) it was found that repetitive elements make up 36.13% of the assembly (more than 80% being unclassified), with the focus being mostly on transposable elements (TEs).

Recent methodological advancements in next-generation sequencing (NGS) and data processing opened up the possibility of obtaining detailed insight into the repetitive composition (repeatome), particularly satDNAs (satellitome) of a genome (Ruiz-Ruano et al. 2016; Pita et al. 2017; Silva et al. 2017). However, until newly arising methods start to be more intensively employed, conventional methods remain the main source of information about tandemly repeated sequences in bivalve genomes.

So far, 48 species belonging to all main bivalve clades have been investigated, yielding 26 different satDNAs, with Mytilidae, Ostreidae and Veneridae being the most explored families so far (Figure 1). The accumulated results point to some peculiarities in bivalve mollusk satDNA structure and organisation, specifically a relatively low genomic content, and extreme evolutionary ancestry and dispersal of some of them, as well as linkage with certain types of mobile elements. In order to summarise the results obtained until now, we present an overview of the most prominent observations on content, features, genomic localisation and species distribution of satDNAs in bivalve mollusk species. Their main characteristics can also be found in Table I.

SatDNAs in Pectinidae

The Antarctic scallop Adamussium colbecki (Smith, 1902) was found to contain a tandemly repeated sequence, detected upon restriction with the
endonuclease BglII (Canapa et al. 2000). The monomer length of this satellite DNA, named pACS, is ~170 bp. Its AT composition ranges from 59.2 to 73%, with the presence of adenine and thymine stretches in the monomer sequence. The degree of similarity between monomers varies between 57 and 95%, and this satellite constitutes about 0.2% of the total DNA of A. colbecki. Sequence motifs similar to CENP-B box and CDEIII were found within the monomer sequence, suggesting that A. colbecki satellite DNA possesses some of the features associated with organisation and activity of the centromere. Subsequent research has shown that it is indeed located in the centromeric region of several A. colbecki chromosomes (Odierna et al. 2006).

In the scallop Pecten maximus (Linnaeus, 1758), six different repetitive DNA sequences have been described (Biscotti et al. 2007). Three repetitive DNAs have monomer lengths of about 40 bp (families 2, 3 and 4), while another three have monomer lengths of 156 bp (family 5), 178 bp (family 1) and 326 bp (family 6). Family 1 represents about 0.13% of the genome. Family 5 monomers are characterised by harbouring successive repeats of one base, in particular G, or GG, AA, TT, CC and AGGG motifs, and having a very low copy number, 0.003%. The longest repeat unit, family 6, consists of two consecutive 163-bp sub-repeats, exhibiting 75% similarity to each other. The authors have also found that it possesses a certain level of similarity with the DTE satellite from the species Donax trunculus Linnaeus, 1758 (Plohl & Cornudella 1996; see below). Fluorescent in-situ hybridisation (FISH) has shown that family 2 was concentrated on a single chromosome pair, while family 1 showed many discrete hybridisation sites varying in size, indicating its presence at multiple chromosomal locations. Family 5
shows six to eight major sites with some more diffuse hybridisation signals, in agreement with its lower genomic abundance.

PjHhaI satDNA was described in *Pecten jacobaeus* (Linnaeus, 1758) upon HhaI digestion of genomic DNA (Petraccioli et al. 2015). This satellite is widespread not only among scallops but also in other bivalve species, including *P. maximus*, *A. colbecki*, *Aequipecten opercularis* (Linnaeus, 1758) and *Chamelea gallina* (Linnaeus, 1758). In *P. jacobaeus* the monomer length is 178 bp, and similarity among monomers is between 92.1 and 95.5%. Sequences obtained from other species ranged from 173 to 181 bp, and showed high similarity to those from *P. jacobaeus* (89.3–93.6%). FISH experiments localised the PjHhaI satDNA to restricted regions of all the bivalents of *P. jacobaeus*, while for other bivalve species hybridisation signals on the chromosomes were scarcely visible. Quantitative analyses showed that the PjHhaI satDNA accounts for ~20% of the genome of *P. jacobaeus* and *P. maximus*. In other bivalve species, dot blot hybridisation signals were barely detectable, making it impossible to determine precisely the very small amount of this satellite DNA. Petraccioli et al. (2015) also found that in the genomes of *C. gigas* and *Capitella teleta* Blake, Grassle and Eckelbarger, 2009 sequences similar to the PjHhaI satDNA exist, flanked by TEs.

### Table I. General features of satDNA sequences presented in the text.

|                  | Monomer length (bp) | AT content (%) | Similarity among monomers (%) | Abundance (%) | Conserved motifs | Connection to mobile elements |
|------------------|---------------------|----------------|-------------------------------|---------------|-----------------|-------------------------------|
| **Pectinidae**   |                     |                |                               |               |                 |                               |
| pACS             | ~170                | 59–73          | 57–95                         | 0.2           | +               |                               |
| Family 1         | 178                 |                | 89.9                          | 0.13          |                 |                               |
| Family 2         | ~40                 |                |                               |               |                 |                               |
| Family 3         | ~40                 |                |                               |               |                 |                               |
| Family 4         | ~40                 |                |                               |               |                 |                               |
| Family 5         | 156                 |                |                               | 0.003         |                 |                               |
| Family 6         | 326                 | 75             |                               |               | +               |                               |
| PjHhaI           | 178                 |                |                               | 20            | +               |                               |
| ApaI 1           | 170–175             |                |                               | +             | +               |                               |
| ApaI 2           | 159–162             |                |                               | +             | +               |                               |
| ApaI 3           | 88–167              |                |                               | +             | +               |                               |
| **Mytilidae**    |                     |                |                               |               |                 |                               |
| ApaI 1           | 173                 | 55             | 91                            | 0.10–0.79     | +               | +                             |
| ApaI 2           | 161                 | 65             | 91                            | 0.85–1.66     | +               | +                             |
| ApaI 3           | 166                 | 55             | 91                            | 0.01–0.12     | +               | +                             |
| Mg1              | 169                 |                |                               | 4.1           |                 |                               |
| Mg2              | 260                 |                |                               | 2.7           |                 |                               |
| Mg3              | 70                  |                |                               | 1             |                 |                               |
| **Donacidae**    |                     |                |                               |               |                 |                               |
| DTHS1            | 136–167             | ~55            | ≤95                           | 0.043         |                 |                               |
| DTHS2            | 136–167             | ~55            | ≤95                           | 0.081         |                 |                               |
| DTHS3            | 136–167             | ~55            | ≤95                           | 0.035         |                 |                               |
| DTHS4            | 136–167             | ~55            | ≤95                           | 0.008         |                 |                               |
| DTE              | 155                 | 89             |                               | 0.09–0.23     | +               | +                             |
| DTF1             | 169                 | 62             | ≥83                           | 0.1           | +               |                               |
| DTF2             | 155,169             | 37.5           | 92–100                        | 2             |                 |                               |
| **Mactridae**    |                     |                |                               |               |                 |                               |
| SSU              | 315                 | 56             | 97                            | 1.3–4         | +               |                               |
| **Ostreidae**    |                     |                |                               |               |                 |                               |
| Cg170/HindIII    | 166                 | 60             | 79–94                         | 1–4           | +               |                               |
| BclI             | 150                 | 50             | 90–96.6                       | +             | +               |                               |
| **Veneridae**    |                     |                |                               |               |                 |                               |
| phBglII400      | 400                 | 60             | 98                            |               |                 |                               |
| **Broadly distributed satDNAs** |          |                |                               |               |                 |                               |
| BIV160           | 158–166             | 61             | 46–100 (interspecies)         | 0.1–2         | +               | +                             |
| DTHS3            | 138–164             | 63             | 48–75 (interspecies)          |               | +               |                               |

**SatDNAs in Mytilidae**

Following genomic DNA restriction with ApaI endonuclease, three different satDNAs were initially detected in *Mytilus edulis* Linnaeus, 1758 (Ruiz-Lara...
et al. 1992), and were subsequently found in the related species *M. galloprovincialis*, *M. trossulus* Gould, 1850 and *M. californianus* Conrad, 1837 (Martínez-Lage et al. 2002). Apal satellite sequences named 1, 2 and 3 exhibit monomer lengths of 173, 161 and 166 bp, respectively. All are AT rich (55–65%), and sequence analysis in *M. edulis* showed a high degree of homology between monomers, 91%. Comparison of satellite profiles between the aforementioned species enabled the reconstruction of phylogenetic relationships between them. It was shown that *M. californianus* is the most diversified among them, which is also consistent with morphological observations, as well as with studies based on mitochondrial DNA. The amount of satellite sequences varies between species but satellite 2 is the most represented in all of them, making up as much as 1.66% of the genome, while satellite 3 is least represented (0.01–0.12%). Chromosomal localisation of these satDNAs was performed for *M. edulis*, *M. galloprovincialis* and *M. trossulus*. While satellite 2 exhibits a random distribution in small groups across all chromosomes, satellites 1 and 3 are located in the subtelomeric areas of individual chromosomes (Martínez-Lage et al. 2002). In continuation, these repetitive DNA sequences were analysed in additional species belonging to *Mytilida* (*M. chilensis* Hupé, 1854, *M. coruscus* Gould, 1861, *Perna canaliculus* (Gmelin, 1791), *Aulacomya ater* (Molina, 1782), *Choromytilus chorus* (Molina, 1782), *Septifer virgatus* (Wiegmann, 1837), *Lithophaga lithophaga* (Linnaeus, 1758), *Geukensia demissa* (Dillwyn, 1817)) as well as in a few other bivalves (*Arca noae* (Linnaeus, 1758), *Pecten maximus*, *Mimachlamys varia* (Linnaeus, 1758)) by Martínez-Lage et al. (2005). These results showed that monomer length and sequence similarities correspond to those reported for satellites from the four initially analysed *Mytilus* species (Ruiz-Lara et al. 1992; Martínez-Lage et al. 2002). The genomic abundance of these satellites is significantly lower in the examined species of *Mytilidae*. Nevertheless, sequence alignments revealed that all of these sequences contain regions of similarity to tRNA A and B boxes, and the authors suggest that they could be spread as tRNA-derived pseudogenes (Martínez-Lage et al. 2005).

Another three repetitive sequences, with monomer lengths of 169, 260 and 70 bp, respectively, were reported in *Mytilus galloprovincialis* (Kourtidis et al. 2006). They were named Mg1, Mg2 and Mg3, constituting 4.1, 2.7 and 1% of the *M. galloprovincialis* genome, respectively. The Mg1 repetitive region with its flanking sequences showed significant homology to the CvE, a member of the pearl family of mobile elements described in the oyster *Crassostrea virginica* (Gmelin, 1791) (Gaffney et al. 2003). The central part of these elements is characterised by a short array of tandemly repeated sequences, similar to satellite monomers. Pearl and other comparably structured elements were recently classified as members of Helentrons, a rolling-circle non-autonomous family of TEs (Thomas & Pritham 2015). Being a putative TE, the structure described by Kourtidis et al. (2006) was named MgE. Similarly, Mg2, and Mg3 repeats were found joined to the previously described Apal type 2 repeats (Ruiz-Lara et al. 1992; Martínez-Lage et al. 2002), in structural arrangements similar to MgE. Kourtidis et al. (2006) have reported that searches for conserved motifs, such as CENP-B and the yeast CDEIII sequence, yielded no significant hits. In addition, sequence alignments did not reveal any other conserved segments.

**SatDNAs in Donacidae**

In the wedge clam, *Donax trunculus*, eight satellite DNAs have been described so far, making it, together with *Pecten maximus*, one of the best-explored bivalve species from a repetitive DNA point of view (Figure 1). Four described satellite DNAs of *D. trunculus* were assigned to the DTHS family, detected after the restriction of genomic DNA with HindIII endonuclease (Plohl & Cornudella 1997). Monomer length of satDNAs within this family ranges from 136 to 167 bp. All four satellites show low abundance: DTHS1 makes up 0.043% of the genome, DTHS2 0.081% and DTHS3 0.035%, while DTHS4 constitutes only 0.008% of the genome. Despite the difference in nucleotide sequence, members of the DTHS family have related oligonucleotide motifs repeated up to 15 times within the satellite monomer. The main building block is a TTAGGG motif and its variations. TTAGGG is a characteristic telomeric sequence in vertebrates, and also builds chromosome ends of *D. trunculus* and some other sea invertebrates (Plohl et al. 2002).

DTE satellite DNA with its 155-bp-long monomer was detected after genomic restriction with EcoRV endonuclease (Plohl & Cornudella 1996). This satellite DNA contains two groups of monomers, DTE1 and DTE2, whose nucleotide sequences show 11% divergence and represent subfamilies that can be clearly distinguished based on diagnostic mutations. DTE1 constitutes 0.23% of the genome, and DTE2 0.09%. This satellite DNA shows similarity to the part of the aforementioned satellite DNA found in *P. maximus* (Biscotti et al. 2007) and to the HindIII satellite DNA of oysters
Following HinfI-cleavage of *D. trunculus* genomic DNA, two different satellite DNAs were detected (Petrović & Plohl 2005). They were separated from the same electrophoretic band, with the only common feature being a repeat length of 169 bp. One of the two satellites is DTF1, which is AT rich (62%) and constitutes 0.1% of the *D. trunculus* genome. DTF1 monomers are divided into two subgroups exhibiting 17% nucleotide divergence, while monomers within each subgroup are highly conserved. Nucleotide diversity analysis of the repeats revealed unequal distribution of mutations, where all DTF1 monomers contain a 30-bp segment showing extremely strong sequence conservation, probably as a consequence of functional constraints on segments of this satDNA. A higher order repeat unit, DTRS, also exists, composed of a complete and a shortened DTF1 monomer and the insertion of a 14-bp-long unrelated nucleotide sequence.

A second HinfI-retrieved satDNA is DTF2. In contrast to other *D. trunculus* satellites, it has an unexpectedly high GC nucleotide content in the monomer (62.5%), and shows high abundance, 2%. Another peculiarity of the DTF2 satellite is partial methylation of monomer sequences. Nucleotide sequence analysis showed that DTF2 monomers have low variability, with nucleotide differences uniformly distributed throughout the whole monomer length. In addition to a 169-bp-long repeat unit (DTF2L variant), a shorter variant of 155 bp was found (DTF2S). FISH showed the presence of this satellite on 14 out of 19 *D. trunculus* chromosomes, residing in telomeric and subtelo-meric regions and, interestingly, mostly not coinciding with the prominent GC-rich heterochromatic blocks (Petrović et al. 2009).

**SatDNAs in Ostreidae**

Clabby et al. (1996) discovered the Cg170 satellite DNA of 166 bp monomer length that occupies 1–4% of the *Crassostrea gigas* genome. The nucleotide sequence of this satellite is AT rich (60%) and the analysis showed a high degree of homology between monomers, 79–94%. Subsequent studies have shown that these satellite sequences are located in the centromeric areas of *C. gigas* chromosomes (Wang et al. 2001). The same satDNA was also found after HindIII restriction of genomic DNA, thereafter named HindIII satDNA, and was described in more detail in seven species belonging to the genera *Ostrea* and *Crassostrea* (Figure 1; López-Flores et al. 2004). Authors reported the consensus length of monomers from these species to be ~166 bp. Species-specific differences among repeats were taxonomically informative and enabled the reconstruction of phylogenetic relationships among examined taxa. In addition, two entries in Repbase (database of repetitive elements, Bao et al. 2015): SAT-1_CGi (Jurka 2012) and SATREP_CGi (Clabby et al. 1996), also belong to the same repetitive sequence. The presence of Cg170/HindIII repeats was also reported by Štavić et al. (2016) within structures that were assigned to DNA transposons of the Helitron superfamily, with varying repeat copy numbers in the internal array.

After digestion with BcII enzyme, a 150-bp-long repetitive DNA sequence was found in the genome of the European flat oyster, *Ostrea edulis* Linnaeus, 1758 (López-Flores et al. 2010). Its presence was determined in five other species belonging to the orders of *Ostrea* and *Crassostrea* (Figure 1). Intraspecific variability is between 3.4% in *Crassostrea angulata* (Lamarck, 1819) and 10% in *C. ariakensis* (Fujita, 1913), with interspecific divergence being similar to intraspecific divergence. AT composition of this satDNA is 50%. In-situ hybridisation analysis showed many signals along all *C. angulata* chromosomes, exhibiting interspersed localisation of BcII repeats. Together with this
organisational pattern, sequence blocks similar to box A and B, characteristic for SINE elements, have been found in BclII repeats. The authors did not determine the exact abundance of this repetitive DNA, but they did report faint hybridisation signals in both Southern and dot blot analysis, indicating a low representation of the BclII sequence in the genomes of *Ostrea stentina* Payraudeau, 1826, *C. angulata* and *C. gigas*, with respect to the genome of *O. edulis*.

**SatDNAs in Veneridae**

phBglII400 satDNA was described by Passamonti et al. (1998) in *Ruditapes philippinarum* (Adams & Reeve, 1850). This is an AT-rich satellite DNA and has 400-bp-long monomers composed of two 200-bp subunits. Sequence analysis shows a high degree of homology between monomers (~98%), and FISH analysis showed that these satellite DNAs are found mainly in pericentromeric chromosomal regions. Its presence was also detected in the species *Venerupis aerea* (Gmelin, 1791) and *Paphia undulata* (Born, 1778).

**SatDNA families broadly distributed among bivalve mollusks**

Although broad interspecies distribution was observed for different mollusk satDNAs (for example, satellites in Pectinidae and Mytilidae, Figure 1), particularly interesting in this context is the BIV160 satDNA. It was initially detected in nine species (*Glycymeris glycymeris* (Linnaeus, 1758), *Mya arenaria* (Linnaeus, 1758), *Dosinia exoleta* (Linnaeus, 1758), *Venus verrucosa* (Linnaeus, 1758), *Venerupis pullastra* (Montagu, 1803), *Venerupis rhomboids* (Pennant, 1777), *Ruditapes decussatus* (Linnaeus, 1758), *Ruditapes philippinarum* and *Nucula* sp. Lamarck, 1799), belonging to all of the main sub-classes of bivalves (Protobranchia, Pteriomorphia and Heteroconchia). According to its distribution, it was estimated to be about 540 million years old, with the longest evolutionary track record known so far (Plohl et al. 2010). In the first survey it remained undetected in *D. trunculus*, but its presence in this species was subsequently confirmed by using specific primers (Šatović et al. 2016). The average AT composition of BIV160 satDNA is 61%, and the monomer length ranges from 158 to 166 bp. Isolated monomers showed three different distribution patterns: species-specific sets of variants (*M. arenaria*), intermixed variants without any species specificity (for example in *R. decussatus*), and monomer variants which can follow both the first and the second pattern (*R. philippinarum*). BIV160 is the most abundant in *R. decussatus*, constituting 2% of the genome, with similarity among monomers ranging from 69 to 97%. Within BIV160 monomers two conserved motifs have been noted, which are also shared with the HindIII satDNA (López-Flores et al. 2004), with DTE (Plohl & Cornudella 1996) and with the CvA mobile element (Gaffney et al. 2003). The similarity of BIV160 sequences and *Grassostrea gigas* Cg170/HindIII repeats (Clabby et al. 1996; López-Flores et al. 2004), the DTE satellite (Plohl & Cornudella 1996), and the CvA element (Gaffney et al. 2003) indicate extreme inter-connections between satDNAs in bivalve species, both in the evolutionary sense and in potential functional roles, particularly with regard to conserved motifs retained in all of them.

The previously mentioned DTHS3 satDNA, originally discovered in *D. trunculus* (Plohl & Cornudella 1997), was further characterised within the class Bivalvia (Šatović & Plohl in press). Monomer variants of DTHS3 satDNA were compared in 12 clam species belonging to subclasses Heterodonta and Pteriomorphia (*D. trunculus*, *Ruditapes decussatus*, *R. philippinarum*, *Sinonovacula constricta* (Lamarck, 1818), *Mercenaria mercenaria* (Linnaeus, 1758), *Meretrix lusoria* (Röding, 1798), *M. meretrix* (Linnaeus, 1758), *Spisula solidissima* (Dillwyn, 1817), *Grassostrea gigas*, *C. virginica*, *Mytilus galloprovincialis* and *Dreissena bugensis* (Andrusov, 1897)). The results suggest that DTHS3 and BIV160 behave quite similarly, both in estimated age (516 MY in case of DTHS3) and in the presence of distribution patterns of monomer variants, which can be species specific, intermingled interspecifically, or a combination of these two patterns. Sequence comparisons also invoked horizontal transfer as a possible pathway by which DTHS3 monomers may become shared between some distant species belonging to different subclasses (Šatović & Plohl In press).

**Integrating features common for bivalve mollusk satDNAs**

One striking characteristic of satellite DNAs present in bivalve genomes is their relatively low genomic abundance, compared to that found in many other organisms (Garrido-Ramos 2017). With the strong exception of the PjHhaI satDNA, which constitutes 20% of the *P. jacobaeus* and *P. maximus* genomes (Petraccioli et al. 2015), the maximum genome occupancy by a certain satDNA is 1–4% for Cg170 in *C. gigas*, being significantly lower for other satellites. What also characterises the satDNA profile of
bivalve mollusks is the coexistence of many divergent subgroups, often interrelated, with none of them expanding into a dominant satDNA of the respective species. An example is *D. trunculus* where eight different satDNAs have been detected, none of them exceeding 2% (Petrović et al. 2009; Plohl et al. 2010), or a family of six repeats detected in the scallop *P. maximus* (Biscotti et al. 2007).

Regarding AT content, bivalves show only a few exceptions to the usual abundance of those nucleotides in the monomer sequence. BclI repeats of *O. edulis* (López-Flores et al. 2010) harbour 50% AT, while DTF2 of *D. trunculus* shows stronger GC enrichment, 62.5% (Petrović et al. 2009).

SatDNA physical mapping in bivalve mollusks encompasses all possible distribution patterns. The presence of BglII satDNA of *A. colbecki* and Cg170 of *C. gigas* was found at centromeres (Canapa et al. 2000; Wang et al. 2001; Odierna et al. 2006). *Mytilus* ApaI satellite 2 shows a random distribution across all chromosomes, while satellites 1 and 3 are located in the subtelomeric areas (Martínez-Lage et al. 2002). Family 2 of *P. maximus* repeats (Biscotti et al. 2007) is present on a single chromosome pair, while family 1 shows many hybridisation signals at multiple chromosomal locations. An interspersed pattern is also characteristic of BclI repeats of *C. angulata* (López-Flores et al. 2010). The distribution pattern of satDNAs is probably correlated with the mechanism of their dispersal; in other words, sequences associated with mobile elements or having the ability to transpose might be expected to build arrays dispersed on many locations along the chromosomes. Although this hypothesis still has to be confirmed, there are indications of such connections, for example in *C. gigas* (Šatović et al. 2016).

Conserved boxes and different sequence motifs can be found within monomer sequences of bivalve satDNAs, and potential functional constraints have already been mentioned.

It is interesting that different types of repetitive sequences, three belonging to satDNAs – DTE (Plohl & Cornudella 1996), BIV160 (Plohl et al. 2010), HindIII (López-Flores et al. 2004) – and one belonging to the mobile element CvA (Gaffney et al. 2003), share the same conserved boxes, although functional studies that would suggest involvement in similar functional interactions do not exist yet. In this respect, it is no surprise that pACS satellite DNA, which contains motifs similar to the CENP-B box and CDEIII (Canapa et al. 2000), was found localised at the centromeres of *A. colbecki* chromosomes (Odierna et al. 2006). In addition to a putative centromeric function, the CENP-B box-like motif could be related to the mobility of satDNA monomers (Meštrović et al. 2013), based on the similarity between the CENP-B protein and transposases of the *pogo* family of mobile elements (Kipling & Warburton 1997).

Sequence alignments have revealed that ApaI-repeats of *Mytilus* species contain segments showing similarity to RNA Pol III boxes A and B (Martínez-Lage et al. 2005), features characteristic of SINE elements. Degenerate sequences similar to boxes A and B have been found within BclI repeats as well (López-Flores et al. 2010). Because the BclI sequence of *O. edulis* exhibits a length and interspersed distribution pattern which is expected for SINE elements, it is likely that the conserved boxes in those repeats could be connected to sequence origin and not a consequence of functional interactions.

The connection of these sequences to mobile elements goes even further, and might suggest a tight interconnection particularly discernible in bivalves, although different forms of interconnection were observed in other species as well (for a review, see Meštrović et al. 2015). As described above, sequence similarity expressed especially through shared conserved boxes has been observed for the CvA element of the pearl family (Gaffney et al. 2003) and DTE, BIV160 and HindIII/Cg170 repeats (Clabby et al. 1996; Plohl & Cornudella 1996; López-Flores et al. 2004; Plohl et al. 2010). *Crassostrea gigas* HindIII/Cg170 repeats together with their flanking sequences were identified as parts of putative TEs (Šatović et al. 2016). MgE, the putative mobile element characterised in *M. galloprovincialis* holds Mg1 repeats (Kourtidis et al. 2006), while sequences showing similarity to P*HhaI* sat in *C. gigas* and *Capitella teleta* were also found surrounded with structures belonging to TEs (Petraccioli et al. 2015). Additionally, DTHS3 satDNA in *M. mercenaria*, *S. solidissima*, *C. virginica* and *M. galloprovincialis* was found embedded in structures that might be responsible for their mobility (Šatović & Plohl in press).

Although scarce and far from being understood, results concerning methylation profiles in repetitive DNAs of bivalve species have disclosed some distinct features. In the oyster *C. gigas*, repeat classes such as DNA transposons, helitrons, satellites, simple repeats and tandem repeats all displayed average methylation levels 2 times higher than the genome background (Wang et al. 2014). These authors noticed that if they divide the repeats into methylated and unmethylated, the divergence rate of the methylated group was significantly lower than that of the unmethylated group. This observation is also in agreement with
data from SSU satDNA from *S. subtruncata*, whose methylation is inversely correlated with nucleotide diversity in segments of monomer sequences (García-Souto et al. 2017). Methylated DTF2 satDNA of *D. trunculus* shows uniform and high sequence conservation throughout the whole monomer sequence (Petrović et al. 2009).

Specificities in the methylation pattern of certain types of repetitive sequences in bivalves suggest role(s) alternative to mere DNA silencing. Furthermore, the existence of a single chromosome pair in *S. subtruncata* in which SSUsat is non-methylated (García-Souto et al. 2017) leads to the assumption that methylation processes in the repetitive part of bivalve mollusk genomes are non-random and complex, and the significance of these patterns remains to be revealed.

Concluding remarks

Research on bivalve mollusk satDNAs has brought important information which has broadened our general views on satDNAs as a class of ubiquitous genomic sequences. For bivalve mollusks, the most striking features are (i) the low contribution of heterochromatin and satDNAs in the genomes of these ancient organisms, including in the centromeric region of some species; (ii) in some cases, an extremely large palette of related monomer variants comprising a library of a satDNA family; (iii) the extremely long ancestry of some of them; (iv) the ability to easily link many of these sequences with certain groups of mobile elements; and (v) unusual methylation patterns. Despite the generally known genomic roles of satDNAs, at this moment it is not clear whether and how specific features of these sequences present in bivalves influence the biology of these organisms. Nevertheless, abundance, dynamics and differences between satDNAs make bivalve mollusks important organisms in exploring organisational and functional aspects of satellites. We are sure that future research will bring a lot of additional information that will help to unravel roles of this specific form of noncoding DNA sequences.

Acknowledgements

We thank Dr Mary Sopka for critical reading and language editing of the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.
Satellite DNA in bivalves

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. Journal of Molecular Evolution 44:189–198. DOI: 10.1007/BF00006135.

Plohl M, Luchetti A, Meštrović N, Mantovani B. 2008. Satellite DNAs between selfishness and functionality: Structure, genomics and evolution of tandem repeats in centromeric (hetero) chromatin. Gene 409:72–82. DOI: 10.1016/j.gene.2007.11.013.

Plohl M, Meštrović N, Mravinac B. 2012. Satellite DNA evolution. In: Garrido-Ramos MA, editor. Genome dynamics. Vol. 7: Basel: Karger. pp. 126–152. DOI: 10.1159/000337122.

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. DOI: 10.1038/hdy.2009.141.

Plohl M, Prats E, Martínez-Lage A, González-Tizon A, Mendez J, Cornudella L. 2002. Telomeric localization of the vertebrate-type hexamer repeat, (TTAGGG)n, in the wedgeshell clam Donax trunculus and other marine invertebrate genomes. The Journal of Biological Chemistry 277:19839–19846. DOI: 10.1074/jbc.M201032200.

Ruiz-Lara S, Prats E, Sainz J, Cornudella L. 1992. Cloning and characterization of a highly conserved satellite DNA from the mollusc Mytilus edulis. Gene 117:237–242.

Ruiz-Ruano FJ, López-León MD, Cabrero J, Camacho JPM. 2016. High-throughput analysis of the satellitome illuminates satellite DNA evolution. Scientific Reports 6:28333. DOI: 10.1038/srep28333.

Saavedra C, Bachère E. 2006. Bivalve genomics. Aquaculture 256:1–14. DOI: 10.1016/j.aquaculture.2006.02.023.

Šatović E, Plohl M. In press. Distribution of DTHS3 satellite DNA across 12 bivalve species. Journal Of Genomics.

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. DOI: 10.1186/s12864-016-3347-1.

Silva DMZDA, Utsunomia R, Ruiz-Ruano FJ, Daniel SN, Porto-Foresti F, Hashimoto DT, Oliveira C, Porto-Foresti F, Foresti F. 2017. High-throughput analysis unveils a highly shared satellite DNA library among three species of fish genus Astyanax. Scientific Reports 7:12726. DOI: 10.1038/s41598-017-12939-7.

Takeuchi T, Kawashima T, Koyanagi R, Gyoda F, Tanaka M, Ikuta T, et al. 2012. Draft genome of the pearl oyster Pinctada fucata: A platform for understanding bivalve biology. DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes 19:117–130. DOI: 10.1093/dnares/dss005.

Thomas J, Pritham EJ. 2015. Helitrins, the Eukaryotic rolling-circle transposable elements. Microbiology Spectrum 3:1–32. DOI: 10.1128/microbiolspec.MDNA3-0049-2014.

Wang X, Li Q, Lian J, Li L, Jin L, Cai H, Xu F, Qi H, Zhang L, Wu F, Meng J, Que H, Fang X, Guo X, Zhang G. 2014. Genome-wide and single-base resolution DNA methylation of the Pacific oyster Crassostrea gigas provide insight into the evolution of invertebrate CpG methylation. BMC Genomics 15:1119. DOI: 10.1186/1471-2164-15-1119.

Wang Y, Xu Z, Guo X. 2001. A centromeric satellite sequence in the Pacific oyster (Crassostrea gigas Thunberg) identified by fluorescence in situ hybridization. Marine Biotechnology 3:486–492. DOI: 10.1016/S1072-0614(00)00106-3.

Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, et al. 2012. The oyster genome reveals stress adaptation and complexity of shell formation. Nature 490:49–54. DOI: 10.1038/nature11413.