

ORIGINAL ARTICLE

Long-term conservation vs high sequence divergence: the case of an extraordinarily old satellite DNA in bivalve mollusks

M Plohl¹, V Petrović¹, A Luchetti², A Ricci², E Šatović¹, M Passamonti² and B Mantovani²

¹Department of Molecular Biology, Ruđer Bošković Institute, Zagreb, Croatia and ²Dipartimento di Biologia Evoluzionistica Sperimentale, Università degli Studi di Bologna, Bologna, Italy

The ubiquity of satellite DNA (satDNA) sequences has raised much controversy over the abundance of divergent monomer variants and the long-time nucleotide sequence stability observed for many satDNA families. In this work, we describe the satDNA BIV160, characterized in nine species of the three main bivalve clades (Protobranchia, Pteriomorpha and Heteroconchia). BIV160 monomers are similar in repeat size and nucleotide sequence to satDNAs described earlier in oysters and in the clam *Donax trunculus*. The broad distribution of BIV160 satDNA indicates that similar variants existed in the ancestral bivalve species that lived about 540 million years ago; this makes BIV160 the most ancient satDNA described so far. In the species examined, monomer variants are distributed in quite a complex pattern. This pattern includes (i) species characterized by a specific group of variants, (ii) species that share distinct group(s) of variants

and (iii) species with both specific and shared types. The evolutionary scenario suggested by these data reconciles sequence uniformity in homogenization-maintained satDNA arrays with the genomic richness of divergent monomer variants formed by diversification of the same ancestral satDNA sequence. Diversified repeats can continue to evolve in a non-concerted manner and behave as independent amplification-contraction units in the framework of a 'library of satDNA variants' representing a permanent source of monomers that can be amplified into novel homogeneous satDNA arrays. On the whole, diversification of satDNA monomers and copy number fluctuations provide a highly dynamic genomic environment able to form and displace satDNA sequence variants rapidly in evolution.

Heredity (2010) **104**, 543–551; doi:10.1038/hdy.2009.141; published online 21 October 2009

Keywords: bivalvia; concerted evolution; conserved/variable boxes; library model; satellite DNA

Introduction

A significant fraction of the eukaryotic genome is made up of satellite DNA (satDNA). These are tandemly repeated sequences that underlie heterochromatic genome compartments (Charlesworth *et al.*, 1994; Schmidt and Heslop-Harrison, 1998). SatDNAs evolve after a pattern known as concerted evolution. Concerted evolution is achieved by the spread or elimination of mutated repetitive units, a process known as homogenization (through mechanisms of non-reciprocal transfer such as gene conversion, unequal crossing-over, rolling circle replication and reinsertion, and transposon-mediated exchange) and variant fixation among reproductively linked individuals (Dover, 1986, 2002). Large differences in satDNA sequences observed among related species are traditionally interpreted as a consequence of accumulated changes in separate lineages in the absence of selective pressure. The library model explains the origin of these differences by amplification-contraction events in a set of satDNAs shared by a group of organisms, a phenomenon that does not necessarily include rapid sequence alterations (Fry and

Salser, 1977; Meštrović *et al.*, 1998, 2006a; Ugarković and Plohl, 2002). In this way, one or a few satDNAs can become highly represented in a taxon, whereas others remain as low-copy number repeats, often not easily detected.

Recent analysis of interspecifically conserved repeats, coexisting in the root-knot nematodes of the genus *Meloidogyne*, suggested two phases in the dynamics of satDNA evolution (Meštrović *et al.*, 2006a). The first phase is tandem amplification of a sequence segment and establishment of a variability profile in the satDNA family. In this phase, sequence spreading can be driven by interplay between selective pressure and stochastic events. The second phase is long-time persistence of satDNA arrays in a library, characterized by reduced homogenization of sequence changes and random-copy number fluctuations.

SatDNAs have been extensively studied in groups such as arthropods and vertebrates, whereas substantial information about the distribution and evolution of this kind of sequences in molluscs is still lacking. Studies performed until now have indicated that bivalve species exhibit a number of different satDNAs in the genome, and some families are shared between species. For example, a recent survey of repetitive DNAs in the king scallop *Pecten maximus* revealed the existence of six different families (Biscotti *et al.*, 2007), one of them exhibiting sequence similarity to the *Donax trunculus* EcoRV (DTE) satellite (Plohl and Cornudella, 1996). A set of three satDNAs has been isolated in mussels of the

Correspondence: Dr M Plohl, Department of Molecular Biology, Ruđer Bošković Institute, Bijenička cesta 54, Zagreb HR-10000, Croatia.
E-mail: plohl@irb.hr

Received 18 March 2009; revised 1 July 2009; accepted 10 September 2009; published online 21 October 2009

genus *Mytilus*: PCR assays have discovered the presence of these satellites in oysters and scallops, but failed to show their occurrence in other Pteriomorphia and in all tested Veneroidea (Martinez-Lage *et al.*, 2005). The clam *Venerupis philippinarum* shares its 400 bp satDNA with some other Tapetinae species, namely *V. aurea* and *Paphia undulata*, but not with *V. decussata* (Passamonti *et al.*, 1998).

In this paper, we characterize a satDNA family related to satellite repeats described in oysters (Lopez-Flores *et al.*, 2004) and in the clam *D. trunculus* (Plohl and Cornudella, 1996). Sequence variants, here grouped as the BIV160 satDNA family, are compared in nine bivalve species of Subclasses Protobranchia, Pteriomorphia and Heteroconchia. The occurrence of BIV160 sequences in various subclasses indicates that this is the oldest satDNA family described so far. The results presented in this paper support the concept of the dual character of satDNAs: the ability of tandem repeats to preserve their nucleotide sequence for long-time periods, and, at the same time, to allow rapid shifts of variants in the repetitive family.

Materials and methods

Samples and DNA extraction

Genomic DNA was isolated from foot muscles of alcohol-preserved specimen of nine species belonging to four orders and three subclasses of Bivalvia using a standard phenol/chloroform protocol (Supplementary material S1). For PCR analyses, DNA was isolated from a single individual of each species, with the exception of *V. decussata*, for which a specimen from each of three different populations has been studied separately (Supplementary material S2). As in the latter case, no differences linked to sampling locality could be observed (not shown), all 37 sequenced monomers were combined in a single set. Large-scale isolation of genomic DNA from *V. decussata* for restriction experiments was carried out from 4 to 5 individuals per population, using the same protocol as above.

PCR amplification, cloning and Southern blot hybridization

All PCR amplifications were performed with the following programme: initial denaturation for 5' at 94°, 35 cycles at 94° for 30''-1', 50°-57° for 30''-1', 72° for 30''-2' and final extension at 72° for 7'. Primers used are listed in Supplementary material S3. PCR on genomic DNA of the crustacean *Triops cancriformis* (Branchiopoda, Notostraca) and a blank reaction were carried out as controls. Amplicons were electrophoresed on a 1.5% agarose gel, DNA in bands of interest eluted using Wizard SV Gel and PCR Clean Up System (Promega, Madison, WI, USA), ligated in a pGEM-T Easy vector (Promega) and used to transform *Escherichia coli* DH5 α competent cells.

Genomic DNA of *V. decussata* was digested with 24 restriction endonucleases (*AccI*, *AluI*, *ApaI*, *AvaI*, *BamHI*, *BclI*, *BglIII*, *DdeI*, *DraI*, *EcoRI*, *HaeIII*, *HindII*, *HindIII*, *HpaI*, *MspI*, *NdeI*, *NsiI*, *PstI*, *RsaI*, *SacI*, *Sall*, *SmaI*, *StuI* and *TaqI*) according to the instructions provided by the manufacturers. *StuI* restriction-obtained 160 bp-long monomer sequences were ligated into a pUC18/*SmaI*-linearized plasmid and transformed as above.

Hybridization was performed using the DIG DNA Labelling and Detection Kit (Roche, Basel, Switzerland),

according to the manufacturer's instructions. DNA probes were digoxigenin labelled and signals detected by CDP-Star (Roche). To assure detection of a broad range of sequences, Southern blots were hybridized at moderate conditions (65 °C; washed at the same temperature in 1 × SSC, 0.1% SDS) with a combination of labelled monomer variants cloned from studied species. Dot blot quantifications were determined by densitometry of hybridization signals of diluted genomic DNA spotted onto a membrane. The calibration curve was constructed from signals of serially diluted cloned fragments.

Sequencing and sequence comparisons

Sequencing was performed at Macrogen Inc. (Korea), with the Big Dye Terminator kit on ABI3730XL DNA Analyzer (Applied Biosystems, Norwalk, CT, USA). Sequences submitted to Genbank obtained the accession numbers EU275729–EU275748 and EU925654–EU925762.

The newly sequenced repeats were analysed together with the following sequences drawn from Genbank: (i) *HindIII* satDNA family from oysters (Lopez-Flores *et al.*, 2004): *Crassostrea gigas* (AJ601431–AJ601437, AJ604547, AJ604548); *C. virginica* (AJ601414–AJ601417); *C. angulata* (AJ601422–AJ601430); *C. gasar* (AJ601418–AJ601421); *C. ariakensis* (AJ604549–AJ604555); *Ostrea edulis* (AJ601406–AJ601413); *O. stentina* (AJ604556–AJ604560); (ii) DTE satDNA (Plohl and Cornudella, 1996), E1 and E2 families (X86926–X86938).

Sequence editing and alignments were performed using the CLUSTAL algorithm of the Sequence Navigator package (Applied Biosystems). BLAST searches (Altschul *et al.*, 1990) were performed with monomer consensus sequences derived for each species. As the boundaries of a satDNA-repeated unit are established arbitrarily, the monomer consensus sequence was duplicated in tandem and the obtained dimeric consensus sequences were used to refine BLAST searches.

Phylogenetic analysis was carried out with the Maximum Parsimony method implemented in PAUP* v40b (Swofford, 2001). All characters were equally weighted. The treatment of gaps either as missing data or as fifth state character did not affect the analysis. Heuristic search was carried out with simple stepwise taxon addition and Tree Bisection-Reconnection branch swapping algorithm. Nodal supports were calculated after 1000 bootstrap replicates, and only clusters with values $\geq 70\%$ were considered as significant.

Analysis of sequence variability

Average monomer length, nucleotide composition and sequence variability, calculated as mean *P*-distances (expressed as percentages) within and between groups, were estimated with MEGA4 software (Tamura *et al.*, 2007). The distribution of nucleotide diversity *P_i* (average number of nucleotide differences per site in a sample population) along the set of satellite monomer variants was calculated using a 10 bp sliding window with a sliding step of 1 bp, using DnaSP v 4.5 (Rozas *et al.*, 2003). Variable and conserved sequence blocks were considered as significant when *P_i* exceeded two times the standard deviation (s.d.) of the average value. Moreover, to make the analysis more stringent, only blocks of a minimum size of a single window (that is 10 consecutive sliding window values exceeding two times s.d.) were used.

Sequence divergence between satDNAs was evaluated through the analysis of transition stages (Strachan *et al.*, 1985). Variations at each nucleotide position in two compared monomer sets are divided in classes from 1 to 6, which follow gradual accumulation of changes in one of the sets at a given position. In this study, original classes have been regrouped as in Meštrović *et al.* (2006a): homogenized stage (class 1), intermediate transition stage (classes 2–4) and advanced stage showing the complete homogeneity of different nucleotides at the same position in each set, and subsequent introduction of a new mutation (classes 5–6). Non-homogenized nucleotide positions occurring concurrently in both sets of monomers could not be grouped according to the method of Strachan *et al.* (1985). However, mutations shared by both groups of variants at a given position are assumed to be inherited from an ancestral set of monomer variants, and are, therefore, considered as a distinct subgroup (see also Mravinac *et al.*, 2005).

Results

Detection and initial characterization of BIV160 satDNA family

In order to search for repetitive sequences related to *Hind*III satDNA from oysters (Lopez-Flores *et al.*, 2004) and DTE satDNA from *D. trunculus* (Plohl and Cornudella, 1996), PCR primers DTE2-78/DTE2-36 were first designed on the basis of alignment of their consensus satellite monomers (Supplementary material S4a). The sampling of repeated sequences through PCR amplification has been used earlier in variability studies of several satellite families. This approach detected a broad range of variants (for example Cesari *et al.*, 2003; Hall *et al.*, 2003), even wider than the set obtained by restriction/cloning (Petrović *et al.*, 2009). PCR products, likely to be specifically amplified from related satDNA variants, were observed in *V. verrucosa* and *V. pullastra* and less defined in *V. decussata*, whereas, in the other tested species, such fragments could not be detected under the reaction conditions used (not shown). Fragments amplified from *V. verrucosa* and *V. pullastra* were cloned and sequenced. Aligned sequences were used to construct new, more specific, primer sets for subsequent screening experiments (VVE-54/VVE-205 and VPU-24/VPU-231;

Supplementary materials S3 and S4b), resulting in amplified tandemly repeated monomer variants of the same satDNA family in all tested species (Supplementary materials S5). As a further test, *V. decussata* genomic DNA was digested with the restriction endonucleases listed in Materials and methods. Digestion with *Stu*I resulted in a ladder-like banding pattern typical for DNA sequences arranged in tandem (Figure 1a), and the circa 160 bp long monomeric restriction fragment was cloned and sequenced. An additional primer set, decStuI-F/decStuI-R, was constructed on the basis of cloned *V. decussata* satDNA monomer (Supplementary materials S3 and S4c), but it turned out to amplify satellite repeats efficiently only in this species and in *Nucula* sp. (see also Table 1). Sequence comparisons of satellite variants obtained either from PCR amplified fragments or from restricted genomic DNA monomers did not reveal differences that would indicate a bias because of the experimental approach. Only fragments larger than the monomer size were used to clone PCR products in order to trim out primer binding sites from sequenced fragments. On the whole, 147 monomers (139 from PCR amplifications and 8 from DNA restriction) were analysed in 9 bivalve species (Table 1).

Average monomer length resides in a narrow range between 158 bp in *V. philippinarum* and 166 bp in *V. verrucosa* (Table 1). The only exception is represented by three monomers in *Glycymeris glycymeris*: these are 278 bp long owing to the internal duplication of a sequence segment within the 159 bp long regular monomer repeat. Nucleotide composition of satellite monomers is biased towards A + T richness, the average content being 61% (Table 1). Alignments revealed sequence similarity between all monomers and allowed derivation of the consensus sequence (Figure 1b). The whole family of monomer variants is henceforth named as BIV160.

BLAST searches of data base entries (June 2009) with either monomeric or dimeric BIV160 consensus sequence revealed expected similarities with sequences of the *Hind*III satDNA family from oysters (Lopez-Flores *et al.*, 2004) and DTE satDNA from *D. trunculus* (Plohl and Cornudella, 1996; Figure 1b). In addition to this, similarity to several entries representing gene flanking sequences or microsatellite-related regions has been found (Supplementary materials S6). The significance of dispersion of BIV160-related monomers or their

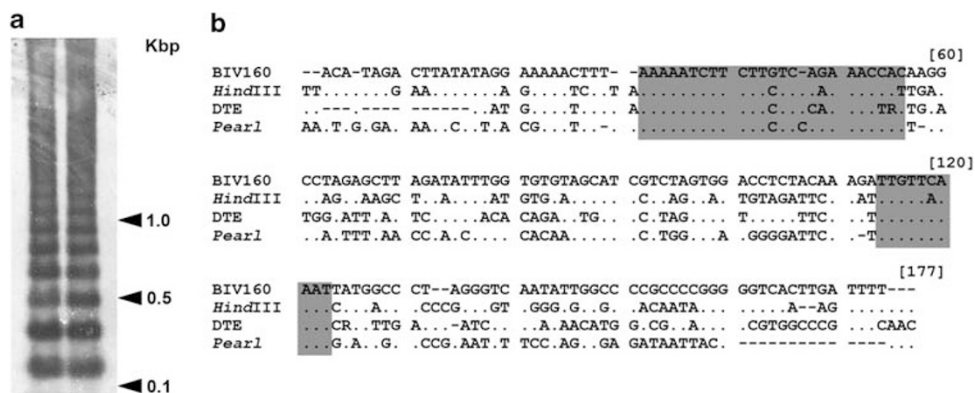


Figure 1 (a) Southern blot analysis of *Stu*I genomic restriction of two *V. decussata* samples (Tunis on the left and Northern Italy on the right). Size markers (kbp) are on the right side. (b) Alignment of monomer consensus sequences derived from BIV160 (this work), *Hind*III (Lopez-Flores *et al.*, 2004) and DTE satDNAs (Plohl and Cornudella, 1996), and MITE-like element *Pearl* (from nucleotide 98–253 of the 465 bp long full size element; Gaffney *et al.*, 2003). Grey boxes highlight regions apparently conserved among all four groups of sequences.

Table 1 Analysed species, with number of sequenced repeats (*N*), respective isolation method, average repeat length, A+T richness, sequence diversity and copy number estimates

Subclass	Order	Species	N	Isolation method ^a	Average repeat length (bp)	A+T (%)	Average sequence variability (%)	Abundance (%)
Protobranchia	Nuculoidea	<i>Nucula</i> sp.	9	3	163	60.2	15.2	n.d.
Pteriomorphia	Arcoida	<i>Glycymeris glycymeris</i>	15	1	159/278 ^b	60.3	25.5	0.01
Heteroconchia	Myoidea	<i>Mya arenaria</i>	6	1	165	59.4	7.9	0.05
		<i>Dosinia exoleta</i>	11	1	160	61.2	22.3	0.05
	Veneroidea	<i>Venus verrucosa</i>	14	1	166	59.3	16.3	0.1
		<i>Venerupis pullastra</i>	25	1,2	162	59.3	16.9	0.1
		<i>Venerupis rhomboides</i>	14	1	163	63.2	26.6	0.02
		<i>Venerupis decussata</i>	37	1,2,3,4	164	60.4	17.3	2
		<i>Venerupis philippinarum</i>	16	1	158	65.6	25.9	0.03

^a1, VVE-54/VVE-205; 2, VPU-24/VPU-231; 3, decStuIF/decStuIR; 4, *StuI* genomic restriction.

^bThree longer monomers were found.

fragments merits further investigation, which will be presented elsewhere, though it may merely reflect the ability of these sequences to move (or to be moved) across genomic locations.

The relative contribution of BIV160 satDNA to the genome of each species (with the exception of *Nucula* sp., owing to the limited quantity of DNA available) was assessed by dot blot analysis (Table 1). Obtained values show that BIV160 is present as a low-copy satDNA (0.01–0.1% of genomic DNA) in all examined species except in *V. decussata*, in which it is represented in significantly higher amounts (2%). It is also possible that the genomic abundance of BIV160 repeats is underestimated because of high diversity of monomer sequence variants. For comparison, E1 and E2 families of *D. trunculus* comprise 0.09 and 0.23% of the genome, respectively (Plohl and Cornudella, 1996), whereas the only available estimate of *HindIII* satDNA in oysters relates to *C. gigas*, whose genome harbours from 1 to 4% of this tandem repeat (Clabby *et al.*, 1996).

Sequence divergence and distribution of mutations in sequenced monomers

For the taxa sequenced in this paper, the intraspecies sequence variability ranges from 7.9% in *Mya arenaria* to 26.6% in *V. rhomboides* (Table 1), whereas the interspecies values range from 16.6% (*Nucula* sp. vs *V. decussata*) to 41.6% (*V. philippinarum* vs *V. rhomboides*). The overall extent of monomer sequence divergence, irrespective of the taxon, is quite large, as it spans the range from complete identity to a divergence of 54.3% (average = 26%). When including *HindIII* monomers of oyster (Lopez-Flores *et al.*, 2004) and *D. trunculus* DTE satellite (Plohl and Cornudella, 1996), the upper limit of this range rises to 65% nucleotide substitutions (average = 35.1%).

The existence of related variants of BIV160 monomers allows the study of sequence variability distribution within and among examined taxa. Sliding window analysis identified both conserved and variable boxes in the majority of species (Figure 2). Each species clearly exhibits its own profile, and a common pattern cannot be derived even when congeneric species of the genus *Venerupis* are considered. In addition to comparisons made after grouping sequences by species of origin, monomers were also analysed by grouping on the basis of the clustering of variants (see next section). Even in

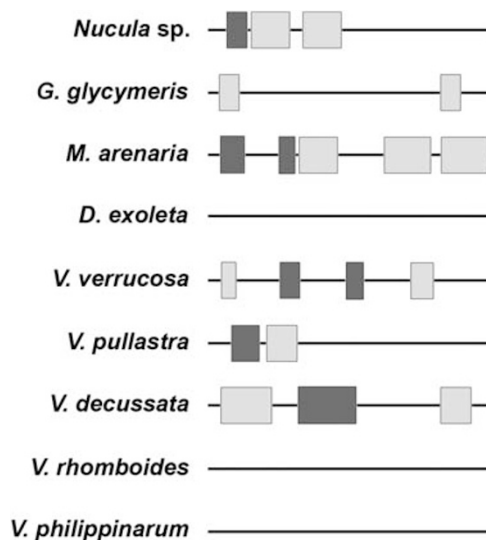


Figure 2 Distribution of nucleotide sequence variability (Π) across repeats. Horizontal lines represent aligned monomer sequences, light and dark grey boxes indicate the positions of conserved and variable regions, respectively.

this case, no pattern of conserved/variable boxes could be retrieved among sequences (not shown). However, visual comparisons of species-defined consensus sequences retrieved two regions of lower variability, constituted by 25 and 10 nucleotides, respectively, that are also shared by *HindIII*, DTE consensus sequences and by the *Pearl* sequence (Figure 1b). Sequence inspection of the two stretches did not reveal any similarity with other known sequences. However, they may indicate a putative motif retained throughout the evolution of these repeats.

Clustering of BIV160 monomer variants

In the Maximum Parsimony heuristic search, the MAXTREE limit was reached at 9900th shortest trees. In the cladogram (Figure 3), three clustering patterns are distinctive. The first concerns species or species-group-specific clusters, as observed for *M. arenaria*, *D. trunculus* and oyster monomers (for clustering in oysters, see Lopez-Flores *et al.*, 2004). The second pattern is just the opposite: repeats cannot be grouped into distinct clusters and cannot be separated into species-specific terminal

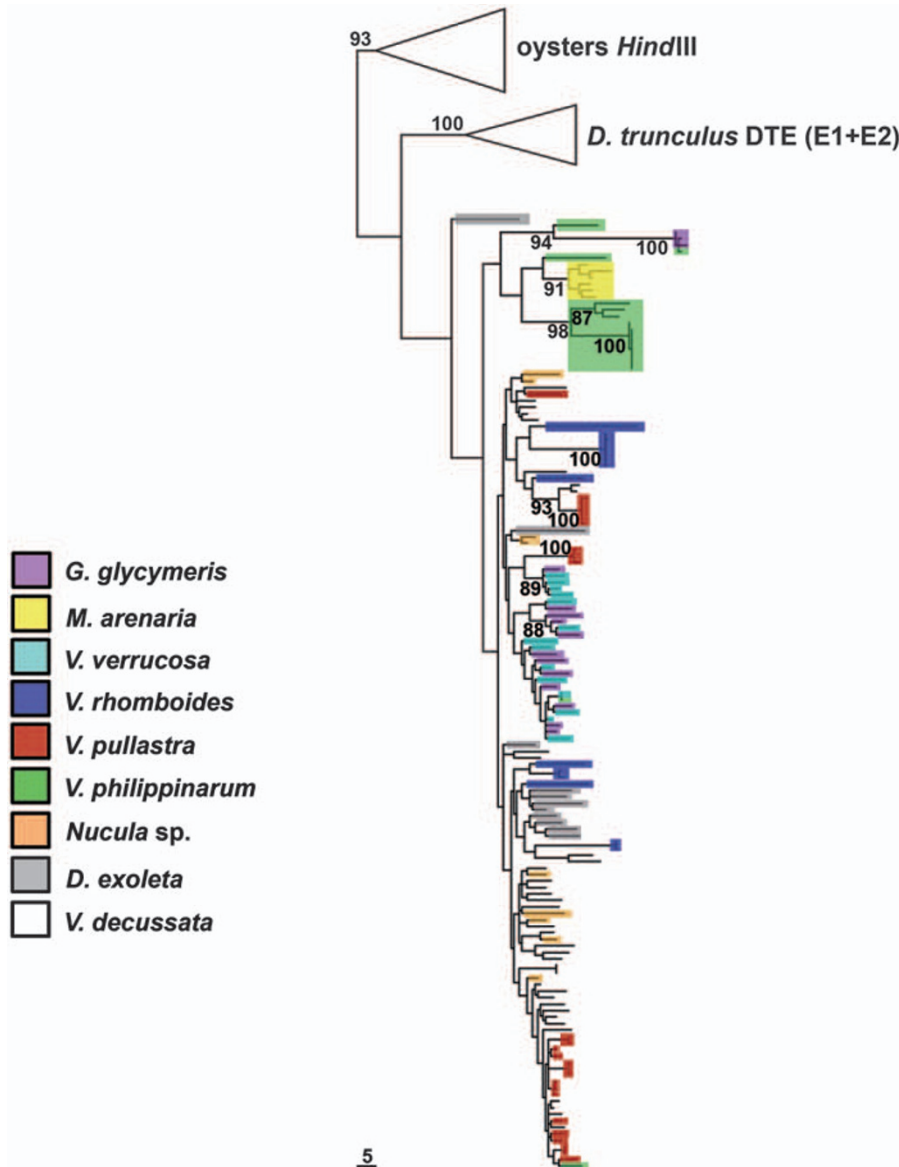


Figure 3 Maximum Parsimony bootstrap consensus tree built on 147 BIV160 monomers sequenced in this work, oyster *HindIII* monomers (Lopez-Flores et al., 2004) and *D. trunculus* DTE satDNA monomers (Plohl and Cornudella, 1996) (TL = 2025; CI = 0.246). Oysters and *D. trunculus* clusters have been collapsed (triangles); bootstrap values >70% are represented at nodes (terminal branching values have been omitted). The bar below the tree indicates five mutational changes.

branches. In *V. decussata*, which is the species with the highest amount of BIV160 satDNA, the analysis of 37 sequenced monomers shows an intermingling with monomers isolated from other species. No specific clustering into subfamilies emerges from diagnostic mutations. The third clustering pattern can be observed for monomers of *V. philippinarum*, *V. rhomboides* and *V. pullastra*: some of their monomers are scattered across the tree, whereas some others are grouped in species-specific branches. More specifically, *V. philippinarum* sequences form one highly supported cluster (green in Figure 3) in which 11 out of the 16 sequenced monomers can be found. The other five monomers are dispersed throughout the tree, diverging from the main cluster by about 39%. Six out of 14 *V. rhomboides* repeats build a single cluster with very low sequence variability (0.2%), but diverging up to 34.4% from other monomers detected in the same species. *V. pullastra* sequences are

grouped in two separate clusters with 5 and 3 monomers (red clusters with 100% bootstrap supports, Figure 3). Although the first cluster appears closely related to two *V. decussata* monomers (93% bootstrap support), the second appears weakly related to the *G. glycymeris*/*V. verrucosa* group (see below). Other *V. pullastra* sequences diverge from the above-mentioned groups by 25.2–26.6% and appear in the less-resolved bottom part of the tree, together with *Nucula* sp., *D. exoleta* and *V. decussata* monomers and one *V. philippinarum* variant. The three longest *G. glycymeris* monomers are highly homogeneous (sequence diversity of 0.2%) and group in a well-supported cluster together with two *V. philippinarum* repeats. The long *G. glycymeris* form shows 42.8% divergence from the standard-length variants, which group together with *V. verrucosa* sequences in several weakly resolved clusters (Figure 3). The complex pattern of relations between BIV160 monomer variants and the

presence of more than one variant type in a single genome prevents the use of this satDNA sequence as a marker to test established phylogenetic relationships.

Transition stages in the process of sequence homogenization

To analyse the extent of sequence differentiation between species in more detail, the Strachan *et al.* (1985) method, modified as published earlier (Meštrović *et al.*, 2006a; see Materials and methods), was applied to sampled pairs of species (Figure 4 and Supplementary material S7): (a) in venerids, as a comparison with similar analysis in oysters (Lopez-Flores *et al.*, 2004); (b) in *Nucula* sp. vs the venerids *D. exoleta* and *V. decussata*, as they lack private clusters in the phylogenetic analysis; (c) in the pair *G. glycymeris* and *V. verrucosa*, as these species pertain to different orders, but their sequences intermingle in the phylogenetic tree.

The analysis of transition stages performed on venerid samples (Figure 4a and Supplementary material S7) shows that the majority of classifiable nucleotide positions detected in given pairs of satDNA subsets are at an intermediate transition stage having from 8.2 to 37.0% of total changes. On the contrary, in the comparison between *V. verrucosa* and *V. rhomboides* just one mutation is in the advanced stage of sequence homogenization. However, the highest number of mutations observed in all comparisons are actually unclassifiable according to the method used (Strachan *et al.*, 1985). Unclassifiable changes include those shared between the two monomer sets assumed to result from homogenization events that occurred before taxa cladogenesis rather than being a consequence of independent mutational events at the same position (Mravinac *et al.*, 2005). In compared venerid species, interspecifically shared mutations represent a large fraction within the category of unclassi-

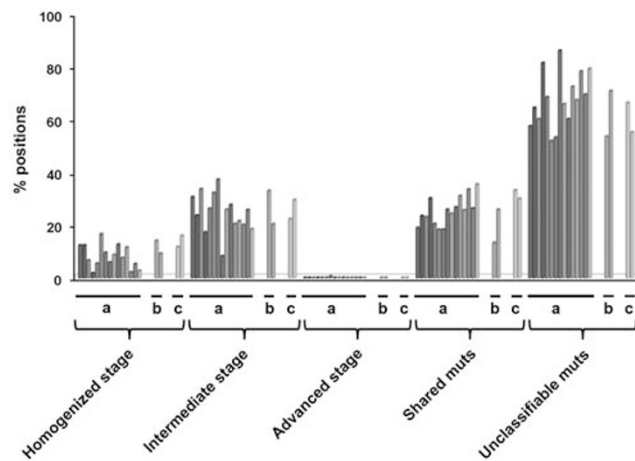


Figure 4 The distribution of transition stages (in percentages) according to the classification described by Strachan *et al.* (1985) and modified in Meštrović *et al.*, 2006a (see explanation in Materials and methods). It should be noted that bars of unclassifiable mutations embody values of changes shared among species and assumed to represent ancestral mutational events. The group of bars 'a' pertains to comparisons between venerids, the group 'b' indicates comparisons between *Nucula* sp. vs *D. exoleta* and *V. decussata* and the group 'c' indicates comparisons between *G. glycymeris* and *V. verrucosa*, considering the set with the longest *G. glycymeris* monomers (left bar) or without them (right bar). For more details on each comparison, refer to Supplementary material S7.

fied changes (from 18.0 to 35.2%), indicating ancient spreading events that occurred before the species split. No detectable shift in the variation profile reflects phylogenetic distances among venerid taxa as was observed in oysters (Lopez-Flores *et al.*, 2004). Here, even when comparisons between more distantly related taxa are taken into account, the picture does not change: comparisons between *Nucula* sp. (Protobranchia) sequences and those from the two venerid taxa *D. exoleta* and *V. decussata* (Heteroconchia) (Figure 4b) or between *G. glycymeris* (Pteriomorphia) and *V. verrucosa* (Heteroconchia) (Figure 4c) show exactly the same pattern.

Discussion

The BIV160 family of satellite repeats is detected in species belonging to the three main clades of the Class Bivalvia (Protobranchia, Pteriomorphia and Heteroconchia). The similarity with *Hind*III satDNA from oysters (Clabby *et al.*, 1996; Lopez-Flores *et al.*, 2004) and DTE satDNA from *D. trunculus* (Pohl and Cornudella, 1996) places all these sequences in the same superfamily of satellite repeats, related in turn to the MITE-like repetitive element *Pearl* found in *Crassostrea* and *Anadara* (Gaffney *et al.*, 2003; Lopez-Flores *et al.*, 2004). The occurrence in species belonging to all three main clades indicates that the origin of these repetitive sequences may be dated back to the origin of the Class Bivalvia itself, deep in the Pre-Cambrian Era (Morton, 1996). In this case, the BIV160 satDNA family may be about 540 million years old, thus representing the oldest described satDNA. Among ancient satDNAs, the *Pst*I family in sturgeons is thought to be >100 million years old (Robles *et al.*, 2004), whereas the identification of sequences related to the primate-specific α -satDNAs in the transcriptome of zebrafish moved the age of this sequence back to the origin of bonefishes, about 400 million years ago (Li and Kirby, 2003). The observed distribution of the three *Mytilus* satellites shared with species from the subclass Pteriomorphia indicates that an expansion should have occurred about 150 million years ago, at the onset of the genus (Martinez-Lage *et al.*, 2005).

The data reported here show the clustering patterns of BIV160 variants isolated from a set of mollusc species. Besides having a broad range of interspecifically intermingled variants, some of the species also gave rise to a few homogeneous monophyletic clusters. Their position within the variability profile of other monomers indicates a recent origin by amplification of a variant from a common pool, probably in the course or after speciation. This distribution is consistent with the satDNA library model (Fry and Salser, 1977; Meštrović *et al.*, 1998, 2006a), but with monomer variants acting as independent amplification-contraction units. Their random amplification in the library (Pons *et al.*, 2004) can explain the dominance of a homogeneous cluster of monomers in *M. arenaria*, the abundance of divergent variants lacking species specificity, or the combination of both features in *V. decussata* and *V. rhomboides*, respectively. A library of satDNA monomer variants has been detected in *Bacillus* stick insects: parthenogenetic taxa lack species-specific variants as a consequence of their kind of reproduction (Cesari *et al.*, 2003).

The overall variability profile of satDNA monomers in a genome is a complex feature that depends on genomic

distribution and homogenization patterns among variants, putative selective constraints imposed on them, as well as on reproduction mode and population factors (Dover, 1986, 2002; Durfy and Willard, 1989; Schindelhauer and Schwarz, 2002; Luchetti *et al.*, 2003; Macas *et al.*, 2006; Meštrović *et al.*, 2006a; Kuhn *et al.*, 2008). Sequence diversity of randomly sampled monomers of a satDNA family in a species usually does not exceed a few per cent (for example Kuhn and Sene, 2004). However, high variability characterizes monomers of some satDNAs, such as in alpha-satellite (Rudd and Willard, 2004), in satDNAs of eusocial insects (Lorite *et al.*, 2004; Luchetti *et al.*, 2006) and in presently sampled BIV160 monomers. Taken together, BIV160 monomer variants accumulate mutations that are not spread among them, suggesting a reduced sequence homogenization and non-concerted evolution. Low efficiency of homogenization processes is supported by high number of ancestral changes that were neither homogenized nor eliminated in the extant taxa. The dominance of ancestral shared changes was also detected in several long-time conserved satDNAs characterized by highly abundant, nearly identical monomers (1–2% divergence) organized in large homogeneous arrays (for example Mravinac *et al.*, 2002, 2005; Meštrović *et al.*, 2006b). Consequently, variability profiles of these monomers do not allow linking variants to the species of origin. Although it is difficult to understand the real causes and consequences of this phenomenon, at present there are two plausible explanations: (a) once established, a mutational profile is favoured because of constraints imposed on the sequence in the heterochromatic environment (Ohta and Dover, 1984; Dover, 1987; Mravinac *et al.*, 2002, 2005; Meštrović *et al.*, 2006a; Plohl *et al.*, 2008) and (b) accumulation and spread of new mutations in a repetitive family is slow owing to biological aspects of DNA sequence evolution in these organisms (Luchetti *et al.*, 2003, 2006; Robles *et al.*, 2004).

The search for possible constraints posed on nucleotide sequences revealed two conserved motifs common to the BIV160 consensus sequences, *Hind*III (Lopez-Flores *et al.*, 2004), DTE (Plohl and Cornudella, 1996) families and the *Pearl* sequence (Gaffney *et al.*, 2003). Broad distribution and ancestry may indicate functional constraints imposed on parts of the monomer sequence. Sequence motifs, in addition to other features (for example inverted repeats, clustered A and T nucleotides, sequence induced DNA curvature; reviewed in Plohl *et al.*, 2004) may be relevant as yet uncharacterized sites involved in DNA–protein interactions in heterochromatin (Hall *et al.*, 2003). For example, motifs related to the human centromeric protein recognition site CENP-B box (Masumoto *et al.*, 1989) were identified in many satDNAs, including those from molluscs (Canapa *et al.*, 2000). Alternatively, the suggested function of conserved segments might be as regions of homology necessary for recombinational processes within and/or between satDNA arrays (Hall *et al.*, 2005).

The unifying feature of many satDNAs is repeat length itself, which is suggested to be required for correct nucleosome positioning or for some other architectural feature of heterochromatin (Henikoff *et al.*, 2001). Despite the relatively high sequence diversity of BIV160 monomers, their length is clustered around 160 bp, with very few exceptions. In addition, similarities in the monomer length

could be an outcome of genomic specificities in turnover rates, as predicted in theoretical models (Stephan and Cho, 1994). For example, in the clam *D. trunculus*, all satDNAs detected until now have about 160 bp long monomers, despite differences in sequence, GC composition, abundance and chromosomal location (Petrović *et al.*, 2009).

The cause of diversity and genomic dispersal of BIV160 sequences can be explained by their relation to the MITE-like element of the *Pearl* family, detected in oysters (Gaffney *et al.*, 2003). Although direct evidence is lacking, the function of transposable elements in the origin and spread of satDNAs has been suggested in many instances, such as in *Drosophila* (Miller *et al.*, 2000), cetaceans (Kapitonov *et al.*, 1998) and human beings (Kipling and Warburton, 1997). Among bivalves, satDNAs found in *Mytilus* show a SINE-like structure (Martinez-Lage *et al.*, 2005). At the moment, it is not possible to establish a direct connection among a MITE-like element, tandem repeats and the presence of dispersed BIV160 sequences found on different genomic locations, including in the vicinity of genes. It can be speculated that transposition might account for the wide distribution that BIV160 shows among bivalve species. Further, lateral transfer in animals is believed to be a rare event (Andersson, 2005). As the distribution pattern of BIV160 variants would require multiple concomitant transfers among species, we consider it as a rather unlikely scenario.

Recent study of a satDNA library in root-knot nematodes indicated long-time persistence of variability profiles of satDNA families (Meštrović *et al.*, 2006a). The BIV160 satDNA seems to be in this phase, in which mutations accumulate, but do not spread among monomer variants. This pattern differs from that followed in BIV160-related *Hind*III satDNA in oysters, in which mutations accumulate in a species-distinctive manner and in correlation with species phylogeny (Lopez-Flores *et al.*, 2004). Differences in satDNA sequence dynamics within and among species are not unusual. For example, sequence variability of a satDNA differs between closely related *Vicia* species, being considerably higher in one taxon than in the other (Macas *et al.*, 2006). In satDNA of the *Drosophila buzzatii* species cluster, all envisaged situations can be found: a pool of ancestral repeats is maintained without being homogenized for any particular variant and regardless of the copy number, whereas in some sets concerted evolution either favours an already existing mutational profile or preferentially homogenizes variants that acquired new changes (Kuhn *et al.*, 2008). However, the reasons for a switch between the types of sequence management are not clear. It may be that the *Hind*III satDNA represents a family that descends from a particular variant, derived from the BIV160 repeats by mutations, homogenized and fixed in oysters. In comparison, homogeneous clusters of BIV160 variants detected in some species (for example in *M. arenaria* in Figure 3) may represent an initial stage of evolution of a new satDNA family.

Taken together, the observations reported here and in several recent papers (Luchetti *et al.*, 2006; Meštrović *et al.*, 2006a; Kuhn *et al.*, 2008) allow us to build an integrated view on pathways of diversification of satDNA sequences in some organisms (Figure 5). A new satellite array is formed by amplification of a sequence segment (Figure 5a). Sequence homogeneity among repeating units is maintained by various recombination mechanisms responsible for the phenomenon of

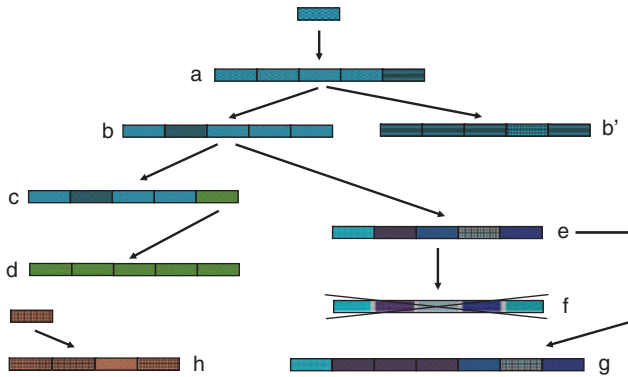


Figure 5 Scenario of diversification of satDNA sequences, unifying long-term sequence stability and potential to generate novel subfamilies through a 'library of variants'. See text for explanation.

concerted evolution (Dover, 1986, 2002), meaning that mutations will either spread among repeat variants in the array (genome) or disappear (Figure 5b). This phase corresponds to the first phase in the life-cycle model of satDNA evolution (Nijman and Lenstra, 2001). Reduced homogenization between variants on different genomic locations can lead to sequence diversification and accumulation of satDNA subfamilies in a genome (Figure 5b and b'; Dover, 1986; Durfy and Willard, 1989; Schindelbauer and Schwarz, 2002). Alternatively, as discussed above, variability profiles can remain 'frozen' during long evolutionary periods (Figure 5c; Mravinac *et al.*, 2002; Meštrović *et al.*, 2006a). Even then, new satDNAs can be formed by amplification of mutated monomers (green box in Figure 5c) evolving in regions of limited sequence homogenization at the array borders (Mravinac and Plohl, 2007, and references therein). Source arrays (Figure 5c) and those formed by mutated monomers (Figure 5d) will presumably evolve independently because of sequence differences among diverged monomers. The resulting diverged repeats can be considered as elements of a satDNA library (Fry and Salser, 1977; Meštrović *et al.*, 1998). An additional source of modified repeats can be in gradually degenerated homogeneous arrays predicted to form the final stage in the satDNA life cycle (Figure 5e; Nijman and Lenstra, 2001). Beside of being lost because of enhanced accumulation of mutations in a non-concerted manner (Figure 5f), the results discussed here suggest that diverged variants can persist in a genome and novel tandem arrays can be formed by amplification of a particular variant (Figure 5g). Together with other DNA sequence(s) expanded into a satDNA (Figure 5h), all variants (Figures 5a–g) contribute to the genomic library of satDNA sequences (Ugarković and Plohl, 2002, Plohl *et al.*, 2008).

The evolutionary pathway presented in Figure 5 unifies two features of satDNAs, homogeneity of arrays and the ability to produce a library of variants derived from the initial sequence, and two patterns, concerted and non-concerted evolution. It must be noted that there is no determinism in this scenario and that re-routings and shortcuts between steps presented here should be possible, resulting in the phenomenon that every diverged variant follows its own path, including being lost. For example, gain and loss of each variant (and/or a satDNA family) may occur at any stage of the cycle because of fluctuations in the copy number. However,

although impossible to predict, gains and losses in the satDNA library can be congruent with species evolutionary history (Meštrović *et al.*, 2009). In this way, satDNAs provide a stable, but at the same time highly flexible system of sequences that, if involved in some functional interactions, can rapidly respond to changes, as proposed for interactions in centromeric regions (Dawe and Henikoff, 2006, Plohl *et al.*, 2008).

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

This work has been supported by Canziani and RFO funds to B Mantovani and M Passamonti and by Research Fund of Republic of Croatia, project no. 098-0982913-2756 to M Plohl. We thank Nevenka Meštrović and Brankica Mravinac for careful reading and useful suggestions given during preparation of this paper, as well as to Mary Sopta for language editing.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *J Mol Biol* **215**: 403–410.
- Andersson JO (2005). Lateral gene transfer in eukaryotes. *Cell Mol Life Sci* **62**: 1182–1197.
- Biscotti MA, Canapa A, Olmo E, Barucca M, Teo CH, Schwarzacher T *et al.* (2007). Repetitive DNA, molecular cytogenetics and genome organization in the King scallop (*Pecten maximus*). *Gene* **406**: 91–98.
- Canapa A, Barucca M, Cerioni PN, Olmo E (2000). A satellite containing CENP-B box-like motifs is present in the Antarctic scallop *Adamussium colbecki*. *Gene* **247**: 175–180.
- Cesari M, Luchetti A, Passamonti M, Scali V, Mantovani B (2003). Polymerase chain reaction amplification of the Bag320 satellite family reveals the ancestral library and past gene conversion events in *Bacillus rossius* (Insecta Phasmatodea). *Gene* **312**: 289–295.
- Charlesworth B, Sniegowski P, Stephan W (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**: 215–220.
- Clabby C, Goswami U, Flavin F, Wilkins NP, Houghton JA, Powell R (1996). Cloning, characterization and chromosomal location of a satellite DNA from the Pacific Oyster, *Crassostrea gigas*. *Gene* **168**: 205–209.
- Dawe RK, Henikoff S (2006). Centromeres put epigenetics in the driver's seat. *Trends Biochem Sci* **31**: 662–669.
- Dover GA (1986). Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. *Trends Genet* **2**: 159–165.
- Dover GA (1987). DNA turnover and the molecular clock. *J Mol Evol* **26**: 47–58.
- Dover GA (2002). Molecular drive. *Trends Genet* **18**: 587–589.
- Durfy SJ, Willard HF (1989). Patterns of intra and interarray sequence variation in alpha satellite from the human X chromosome: evidence for short range homogenization of tandemly repeated DNA sequences. *Genomics* **5**: 810–821.
- Fry K, Salser W (1977). Nucleotide sequences of HS-alpha satellite DNA from kangaroo rat *Dipodomys ordii* and characterization of similar sequences in other rodents. *Cell* **12**: 1069–1084.
- Gaffney PM, Pierce JC, Mackinley AG, Titchen DA, Glenn WK (2003). *Pearl*, a novel family of putative transposable elements in bivalve molluscs. *J Mol Evol* **56**: 308–316.
- Hall SE, Kettler G, Preuss D (2003). Centromere satellites from *Arabidopsis* populations: maintenance of conserved and variable domains. *Genome Res* **13**: 195–205.

- Hall SE, Luo S, Hall AE, Preuss D (2005). Differential rates of local and global homogenization in centromere satellites from *Arabidopsis* relatives. *Genetics* **170**: 1913–1927.
- Henikoff S, Ahmad K, Malik HS (2001). The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* **293**: 1098–1102.
- Kapitonov VV, Holmquist GP, Jurka J (1998). L1 repeat is a basic unit of heterochromatin satellites in cetaceans. *Mol Biol Evol* **15**: 611–612.
- Kipling D, Warburton PE (1997). Centromeres, CENP-B and Tigger too. *Trends Genet* **13**: 141–145.
- Kuhn GCS, Sene FM (2004). Characterization and interpopulation variability of a complex *HpaI* satellite DNA of *Drosophila seriema* (repleta group). *Genetica* **121**: 241–249.
- Kuhn GCS, Sene F, Moreira-Filho O, Schwarzacher T, Heslop-Harrison JS (2008). Sequence analysis, chromosomal distribution and long-range organization show that rapid turnover of new and old pBuM satellite DNA repeats leads to different patterns of variation in seven species of the *Drosophila buzzatii* cluster. *Chromosome Res* **16**: 307–324.
- Li Y-X, Kirby ML (2003). Coordinated and conserved expression of aliphoid repeat and aliphoid repeat-tagged coding sequences. *Dev Dynam* **228**: 72–81.
- Lopez-Flores I, de la Herran R, Garrido-Ramos MA, Boudry P, Ruiz-Rejon C, Ruiz-Rejon M (2004). The molecular phylogeny of oysters based on a satellite DNA related to transposons. *Gene* **339**: 181–188.
- Lorite P, Carrillo JA, Tinaut A, Palomeque T (2004). Evolutionary dynamics of satellite DNA in species of the genus *Formica* (Hymenoptera, Formicidae). *Gene* **332**: 159–168.
- Luchetti A, Cesari M, Carrara G, Cavicchi S, Passamonti M, Scali V et al. (2003). Unisexuality and molecular drive: Bag320 sequence diversity in *Bacillus* taxa (Insecta Phasmatodea). *J Mol Evol* **56**: 587–596.
- Luchetti A, Marini M, Mantovani B (2006). Non-concerted evolution of the RET76 satellite DNA family in *Reticulitermes* taxa (Insecta, Isoptera). *Genetica* **128**: 123–132.
- Macas J, Navratilova A, Koblizkova A (2006). Sequence homogenization and chromosomal localization of VicTR-B satellites differ between closely related *Vicia* species. *Chromosoma* **115**: 437–447.
- Martinez-Lage A, Rodriguez-Farina F, Gonzalez-Tizon A, Mendez J (2005). Origin and evolution of *Mytilus* mussel satellite DNAs. *Genome* **48**: 247–256.
- Masumoto H, Masukata H, Muro Y, Nozaki N, Okazaki T (1989). A human centromere antigen (CENP-B) interacts with a short specific sequence in aliphoid DNA, a human centromere satellite. *J Cell Biol* **109**: 1963–1973.
- Meštrović N, Plohl M, Mravinac B, Ugarković Đ (1998). Evolution of satellite DNAs from the genus *Palorus*—experimental evidence for the library hypothesis. *Mol Biol Evol* **15**: 1062–1068.
- Meštrović N, Castagnone-Sereno P, Plohl M (2006a). 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.
- Meštrović N, Castagnone-Sereno P, Plohl M (2006b). High conservation of the differentially amplified MPA2 satellite DNA family in parthenogenetic root-knot nematodes. *Gene* **376**: 260–267.
- Meštrović N, Plohl M, Castagnone-Sereno P (2009). Relevance of satellite DNA genomic distribution in phylogenetic analysis: a case study with root-knot nematodes of the genus *Meloidogyne*. *Mol Phyl Evol* **50**: 204–208.
- Miller WJ, Nagel A, Bachmann J, Bachmann L (2000). Evolutionary dynamics of the SGM transposon family in the *Drosophila obscura* species group. *Mol Biol Evol* **17**: 1597–1609.
- Morton B (1996). The evolutionary history of Bilvalvia. In: Taylor J (ed). *Origin and Evolutionary Radiation of the Mollusca*. Oxford University Press: Oxford, pp 337–359.
- Mravinac B, Plohl M, Meštrović N, Ugarković Đ (2002). Sequence of PRAT satellite DNA ‘frozen’ in some Coleopteran species. *J Mol Evol* **54**: 774–783.
- Mravinac B, Plohl M, Ugarković Đ (2005). Preservation and high sequence conservation of satellite DNAs suggest functional constraints. *J Mol Evol* **61**: 542–550.
- Mravinac B, Plohl M (2007). Satellite DNA junctions identify the potential origin of new repetitive elements in the beetle *Tribolium madens*. *Gene* **394**: 45–52.
- Nijman IJ, Lenstra JA (2001). Mutation and recombination in cattle satellite DNA: a feedback model for the evolution of satellite repeats. *J Mol Evol* **52**: 361–371.
- Ohta T, Dover GA (1984). The cohesive population genetics of molecular drive. *Genetics* **108**: 501–521.
- Passamonti M, Mantovani B, Scali V (1998). Characterization of a highly repeated DNA family in Tapetinae Species (Mollusca Bivalvia: Veneridae). *Zool Sci* **15**: 599–605.
- Petrović V, Perez-Garcia C, Pasantes JJ, Šatović E, Prats E, Plohl M (2009). A GC-rich satellite DNA and karyology of the bivalve mollusc *Donax trunculus*: a dominance of GC-rich heterochromatin. *Cytogenet Genome Res* **124**: 63–71.
- Plohl M, Cornudella L (1996). Characterization of a complex satellite DNA in the mollusk *Donax trunculus*: analysis of sequence variations and divergence. *Gene* **169**: 157–164.
- Plohl M, Bruvo B, Meštrović N, Mravinac B, Petrović V, Durajlija-Žinić S et al. (2004). Satellite DNA sequences in centromeric heterochromatin. *Periodicum Biol* **106**: 95–102.
- 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.
- Pons J, Bruvo B, Petitpierre E, Plohl M, Ugarković Đ, Juan C (2004). Complex structural features of satellite DNA sequences in the genus *Pimelia* (Coleoptera: Tenebrionidae): random differential amplification from a common ‘satellite DNA library’. *Heredity* **92**: 418–427.
- Robles F, de la Herran R, Ludwig A, Ruiz Rejon C, Ruiz Rejon M, Garrido-Ramos MA (2004). Evolution of ancient satellite DNAs in sturgeon genomes. *Gene* **338**: 133–142.
- Rozas J, Sánchez-Del Barrio JC, Messegyer X, Rozas R (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496–2497.
- Rudd MK, Willard HF (2004). Analysis of the centromeric regions of the human genome assembly. *Trends Genet* **20**: 529–533.
- Schindelhauer D, Schwarz T (2002). Evidence for a fast, intrachromosomal conversion mechanism from mapping of nucleotide variants within a homogeneous α -satellite DNA array. *Genome Res* **12**: 1815–1826.
- Schmidt T, Heslop-Harrison JS (1998). Genomes, genes and junk: the largescale organization of plant chromosomes. *Trends Plant Sci* **3**: 195–199.
- Strachan T, Webb D, Dover GA (1985). Transition stages of molecular drive in multiple-copy DNA families in *Drosophila*. *EMBO J* **4**: 1701–1708.
- Stephan W, Cho S (1994). Possible role of natural selection in the formation of tandem-repetitive noncoding DNA. *Genetics* **136**: 333–341.
- Swofford DL (2001). *PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4b. Sinauer Associates: Sunderland, Massachusetts.
- Tamura K, Dudley J, Nei M, Kumar S (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software. Version 4.0. *Mol Biol Evol* **24**: 1596–1599.
- Ugarković Đ, Plohl M (2002). Variation in satellite DNA profiles-causes and effects. *EMBO J* **21**: 5955–5959.

Supplementary Information accompanies the paper on Heredity website (<http://www.nature.com/hdy>)