

REVIEW

Satellite DNA in insects: a review

T Palomeque and P Lorite

Departamento de Biología Experimental, Área de Genética, Universidad de Jaén, Jaén, Spain

The study of insect satellite DNAs (satDNAs) indicates the evolutionary conservation of certain features despite their sequence heterogeneity. Such features can include total length, monomer length, motifs, particular regions and/or secondary and tertiary structures. satDNAs may act as protein-binding sites, structural domains or sites for epigenetic modifications. The selective constraints in the evolution of satDNAs may be due to the satDNA sequence interaction with specific proteins important in heterochromatin formation and possible a role in controlling gene expression. The transcription of satDNA has been described in vertebrates, invertebrates and plants. In insects, differential satDNA expression has been observed in different cells, develop-

mental stages, sex and caste of the individuals. These transcription differences may suggest their involvement in gene-regulation processes. In addition, the satDNA or its transcripts appear to be involved in heterochromatin formation and in chromatin-elimination processes. The importance of transposable elements to insect satDNA is shown by their presence as a constituent of satDNA in several species of insects (including possible active elements). In addition, they may be involved in the formation of centromeres and telomeres and in the homogenization and expansion of satDNA.

Heredity (2008) **100**, 564–573; doi:10.1038/hdy.2008.24; published online 16 April 2008

Keywords: heterochromatin; evolution of repetitive DNA; CENP-B box; transcription satellite DNA; transposable elements

Introduction

Satellite DNA (satDNA) constitutes a considerable part of the genomic DNA of eukaryotic organisms, being the major DNA component of heterochromatin. It is located mainly in the pericentromeric and/or telomeric regions of chromosomes (Charlesworth *et al.*, 1994). satDNA is generally formed by long tandem arrays in which the monomers (or repeat units) are repeated in a head-to-tail fashion. This molecular organization gives rise to a characteristic ladder of bands (multimeres of a basic satellite-repeat unit), after agarose gel electrophoresis of genomic DNA digested with the appropriate restriction endonucleases. However, this methodology does not provide sufficient information on such aspects as the total length of the satellite arrays, the consecutive order of satellite variants (or subfamilies) and the pattern interruption and features of the sequences inserted within a satellite array. Neither does it enable the detection of satDNA when it is present in very low copy number. PCR techniques using the appropriate primers solve, at least in part, some of these problems.

Recent technological advances have allowed the genome sequence to be determined successfully in a number of eukaryotic organisms including *Drosophila* and, on a smaller scale, other insects (Krzywinski *et al.*,

2005; Hoskins *et al.*, 2007; Schittenhelm *et al.*, 2007; Smith *et al.*, 2007). However, in most cases, only the euchromatic part of the genome has been satisfactorily determined. Large heterochromatic segments remain poorly analyzed, since the repetitive nature of the DNA present in heterochromatic regions makes cloning, assembly and annotation very difficult.

The heterochromatin is important in the establishment and maintenance of the centromeric, telomeric and subtelomeric regions, which are essential for proper chromosome segregation. In addition, the heterochromatin harbors vital genes (Dimitri *et al.*, 2005). Several functional roles have been suggested for satDNA, the major DNA component of heterochromatin, although the details of the molecular mechanisms remain unclear. Noncoding RNAs have been associated with such processes as the maintenance and spreading of silent chromatin, dosage compensation and the programmed DNA elimination that distinguished the germ line from soma in some organisms. The transcripts of the tandem-repeat centromeric DNA of the fission yeast *Schizosaccharomyces pombe* have been found to be clearly involved in RNA interference (RNAi)-mediated heterochromatin assembly. Similarly, recent data on DNA elimination in the ciliated protozoans *Tetrahymena* and *Paramecium* appear to indicate that it occurs via an RNAi-like mechanism (reviewed by Bernstein and Allis, 2005). However, there are many aspects that remain unknown about these processes.

In this review, we summarize the existing knowledge of satDNA in insects, or at least a great part of it. These data have been compiled in Table 1 (Supplementary information). Despite the existence of nearly a million insect species, satDNA been studied in only on a few

Correspondence: Dr T Palomeque, Departamento de Biología Experimental, Área de Genética, Universidad de Jaén, Campus de las Lagunillas sn, Jaén 23071, Spain.

E-mail: tpalome@ujaen.es

Received 15 May 2007; revised 5 February 2008; accepted 16 March 2008; published online 16 April 2008

species, belonging to 8 out of 32 orders. This review summarizes the characteristics of satDNA, some of which may be of functional importance. We suggest that the transcription of satDNA in insects probably occurs in a high number of species, although this remains to be demonstrated.

Characteristics and properties of insect satellite DNA

Sizes of repetitive units and specificity of distribution of the satellite DNA

Insect satDNA has been classified as simple or complex according to the length of the repetitive units (King and Cummings, 1997). For example, most satDNAs from *Drosophila* can be divided into two such groups. One group formed by tandem repeats of a simple sequence of only 5, 7 or 10 bp in length, corresponding to 1.672, 1.686 and 1.705 g cm⁻³ bands in CsCl. The other group is formed by the 1.688 g cm⁻³ satellite, in which the major component is consisted of the 359-bp repeats from the X chromosome, although other satellite variants have also been found (Bonaccorsi and Lohe, 1991 and references therein). The presence of minor variants of the 1.688 satellite in pericentromeric region from other chromosomes has been also detected (Abad *et al.*, 2000). King and Cummings (1997) considered that the satDNA of the remaining insects falls in two size classes, one in the range of about 140–190 bp and the other in the range of about 300–400 bp. The two classes (Table 1 in Supplementary information) can be recognized, although with numerous and remarkable exceptions, such as 24-bp satDNA from *Musca domestica* (Blanchetot, 1991), 44-bp satDNA from *Ceratitis capitata* (Stratikopoulos *et al.*, 2002), 1169-bp satDNA from *Misolampus goudoti* (Pons, 2004) and 2.5-kb satDNA from *Monomorium subopacum* (Lorite *et al.*, 2004a).

In relation to its specific distribution (Table 1 in Supplementary information), satDNA is sometimes species specific, as in the case of the 542-bp satDNA from *Gryllus bimaculatus* (Yoshimura *et al.*, 2006b), whereas other variants are shared among more or less related species, such as the 180-bp satDNA from *Drosophila ambigua*, which is also present in *D. tristis* and *D. obscura* (Bachmann and Sperlich, 1993). It bears mentioning that certain satDNAs that are presently considered species specific may actually be present in other related species, in low copy number detectable only by PCR assays, as discussed below.

Generally the same type of satDNA exists in all chromosomes of an insect species. For example, *Palorus subdepressus* has the same satDNA in the pericentromeric heterochromatic region on all chromosomes (Plohl *et al.*, 1998). The same type of satDNA is also present in all chromosomes of the leaf beetle *Chrysolina americana* (Lorite *et al.*, 2001).

However, sometimes the satDNA can be chromosome specific. The most well-known case is found in *D. melanogaster*, in which each centromeric region has different repeated DNA sequences (Bonaccorsi and Lohe, 1991 and references therein). It can also be sex-chromosome specific. For example, the chromosome X of the aphid genus *Megoura* has a large heterochromatic block with a specific satDNA (Bizzaro *et al.*, 1996). It has been

suggested that these heterochromatic blocks may be involved in the delay of X chromosome separation during the maturation of aphid parthenogenetic oocytes, which is considered the basis of male sex determination, although data are limited on this topic (Mandrioli *et al.*, 1999 and references therein).

Satellite DNA and DNA curvature

One of the most widespread characteristics of satDNA in insects is an intrinsically bent structure (Table 1 in Supplementary information). This characteristic is shared with the satDNA from the majority of eukaryotic organisms. Intrinsic curvature is a sequence-dependent property of the DNA molecule. The curvature pattern of satDNA has been studied by examining its mobility in non-denaturing polyacrylamide gel electrophoresis. DNA curvature and its tertiary structure can be also studied using predictive models of sequence-dependent DNA-bending programs. The richness of A–T as well as the existence of clusters of d (A–T) of ≥ 3 residues periodically spaced along the DNA molecule with a period close to that of the helical repeat has been related to the degree of DNA curvature. The study of satDNA from three subspecies of the beetle *Pimelia sparsa* has shown that the properly phased A-tracts are a fundamental feature of DNA curvature (Barceló *et al.*, 1997). This relationship has also been experimentally supported, for satDNA from the ant *M. subopacum* (Lorite *et al.*, 2004a) and for the curved satDNA of 35 taxa from the beetle genus *Pimelia* (Pons *et al.*, 2004 and references therein). However, it has also been reported that other nucleotide tracts (mainly phase one) are probably involved in the bending of DNA (Carrera and Azorin, 1994; Barceló *et al.*, 1998). Despite the above data, the structural and molecular properties of the DNA involved in the curvature are not yet well known (Matsugami *et al.*, 2006). The potential role of DNA curvature is not well established, but it may be related to chromatin organization and the tight winding of DNA in constitutive heterochromatin as well as to specific protein binding (Lobov *et al.*, 2001).

It has been reported that the inverted and palindromic repeats and especially dyad and cruciform structures could operate as nucleosome-positioning signals as an alternative to the intrinsically curvature of the DNA (Barceló *et al.*, 1998). The satDNAs from the parasitic wasps *Diadromus* and *Eupelmus* have conserved inverted repeats that may adopt secondary dyad structures that may be important in heterochromatin condensation (Rojas-Rousse *et al.*, 1993). Similar structures have been reported for the satDNA of *Trichogramma brassicae* (Landais *et al.*, 2000), *Tribolium* sp. (Mravinac *et al.*, 2005a) and *Chironomus pallidivittatus* (Rosén *et al.*, 2002 and references therein).

Complex repeat organization of the satellite DNA

Some satDNAs show complex repeat organization, resulting in higher-order repeats (HORs). These complex and longer repeats maintain a high sequence similarity (between higher orders but not within them). Within an HOR the monomers can show remarkable sequence divergence. Sometimes even the HORs are composed of different subfamilies as happens, for example, in the

well-known case of human alphoid sequences of chromosome 7 (Willard and Wayne, 1987). Similarly, the satDNA from the cave beetle *Pholeuon proserpinae* has a 532-bp HOR composed of two types of 266-bp monomers. The dimers, invariably composed of a monomer of each type, appear to be the repetitive units that undergo concerted evolution and show high sequence identity as a result of the homogenization process (Pons *et al.*, 2003).

More complex HORs have been found in the satDNA from the phytophagous beetle *Chrysolina carnifex*. These satDNA show six different 211-bp monomer types clearly separated in the phylogenetic tree, although they probably have a common evolutionary origin. They are organized in three types of repeats; monomers (211 bp) and HORs in the form of dimers (477 bp) or even trimers (633 bp). The sequencing of DNA fragments of high molecular weight, fluorescence *in situ* hybridization and Southern hybridizations suggest that each type of repeat is intermixed in the heterochromatic regions (Palomeque *et al.*, 2005).

Other satDNAs also have a complex structure; for example, the 1061-bp satDNA from the beetle *Tribolium brevicornis* has a structure based on two long repeats with about 470 bp (inversely oriented and with a high capacity to form a thermodynamic dyad) and also includes two segments (of 56 and 65 bp) that alternate between the 470-bp repeats. This satDNA could have a complex origin, including the spread of an inversely duplicated element in a HOR with a monomer of about 470 bp (Mravinac *et al.*, 2005a).

Transcription of the satellite DNA

Transcription of satDNA has been reported in vertebrates, invertebrates and plants. The transcription of satDNAs generally shows developmental-stage and tissue-specific differences, suggesting that the transcripts could have regulatory roles, although the molecular mechanism of action is still unknown (reviewed by Ugarkovic, 2005).

RNAi has been related to the recognition of repetitive DNA elements as a preferential target for heterochromatin assembly. In addition to fission yeast, a connection between RNAi and centromeric heterochromatin formation has been described in plants, insects and mammals (reviewed in Bernstein and Allis, 2005). In *Drosophila*, small-interfering RNAs (siRNAs) have been isolated, and these are cognate to several types of repetitive DNA, suggesting that the small RNAs are involved in the process of chromatin modification. These RNAs are most abundant in testes and early embryos. This fact is probably related to the marked and dramatic changes in the heterochromatin structure in these stages (Aravin *et al.*, 2003).

Some siRNAs corresponding to 1.688 satDNA have also been detected (Aravin *et al.*, 2003). Recent studies support the contention that the transcription of this satDNA in ovaries could be under the control of RNAi machinery to maintain the silenced state of these centromeric and pericentromeric repeats (Usakin *et al.*, 2007).

SatDNA transcripts have been related to the introns of dynein-encoding mega-genes on the Y chromosome from *D. melanogaster*, *D. hydei* and *D. eohydei*. These

male-fertility genes on the heterochromatic Y chromosome are characterized by their size—in the range of several megabases—by their expression being limited to premeiotic spermatocytes and generally by their association with enormous species-specific lampbrush loops. Each loop consists of a DNA axis associated with huge species-specific repetitive transcripts and with large amounts of non-Y-encoded proteins. The loop-forming regions consist of species-specific satDNA interspersed with transposable elements (TEs). The transcripts of the simple AAGAA repeats (a component of the 1.686 g cm⁻³ satellite) have been found on the Y chromosome loops in *D. melanogaster* (Bonaccorsi and Lohe, 1991). Recent papers support a coding function of the Y-linked fertility factors. However, it is not clear why giant, lampbrush loops are formed. Nor is the biological significance of their protein-binding function understood (Trapitz *et al.*, 1988; Kurek *et al.*, 2000; Piergentili *et al.*, 2004 and references therein).

The 500-bp satDNA family from the cave cricket *Dolichopoda schiavazzii* is transcribed and the transcripts can function as ribozymes with self-cleavage activity, although their physiological function remains unknown (Rojas *et al.*, 2000 and references therein). Transcription of satDNA has also been reported from several hymenopteran species, including the sawfly *Diprion pini*, the parasitic wasp genera *Diadromus* and *Eupelmus*, the bumblebee *Bombus terrestris* (Rouleux-Bonnin *et al.*, 2004 and references therein) and the ant *Aphaenogaster subterranea* (Lorite *et al.*, 2002b). Generally, the amounts of transcribed satDNA differ among the queen, worker and male genomes. All these transcribed satDNAs are curved or potentially curved. The rate of satDNA transcription differs in female and male species, and a female sex-specific transcript exists. The satDNAs are transcribed on both strands in both sexes, except in *B. terrestris*, where the satDNA is transcribed on both strands in adults but preferentially on one strand in the embryos, suggesting an instar-dependent transcriptional activity and different expression patterns during the differentiation process (Lorite *et al.*, 2002b; Rouleux-Bonnin *et al.*, 2004).

Rouleux-Bonnin *et al.* (2004) suggested that satDNA transcription may be initiated within the satDNA. The authors point out the existence of potential transcription regulatory elements for RNA polymerase II and III in *Diadromus* satDNA. They conducted an exhaustive study of satDNA curvature, showing the different curved states possible; in each state of curvature, the protein interaction could vary. The interplay between HMG-D with histone H1 appears to be important in the chromatin-assembly process during early embryogenesis in *Drosophila*, whereas its absence is correlated with transcriptional competences (Ner *et al.*, 2001). Similar observations have been reported in other organisms (Dimitrov and Wolfe, 1996). The DNA curvature may have two roles, one in compacting the chromatin and the other in changing the amount of satDNA during sexual differentiation and its specific transcription during development. Consequently, it is possible that although the gene *csd* (complementary sex determiner) is the primary signal for sexual development, as reported by Beye *et al.* (2003) in the hymenopteran honeybee, the satDNA may be related to sex- and caste-differentiation processes.

Satellite DNA and chromatin-elimination processes

The DNA elimination in *Tetrahymena* and *Paramecium* appears to occur via an RNAi-like mechanism. There are another two uncharacterized DNA-elimination processes in which satDNA may be involved.

One is seen in the chironomid *Acricotopus lucidus*, where part of the chromosome complement is eliminated from somatic cells during germ line-soma segregation, giving rise to two types of chromosomes: the germ line-limited chromosomes (Ks) and the soma chromosomes (Ss). The Ks consist of large S-homologous sections and of heterochromatic segments with a germ line-specific repetitive DNA family. It has been suggested that these sequences may have importance in the germ line-soma segregation processes, probably by acting as enzymatic signals for the recognition of both lines (Staiber, 2002, 2004, 2006).

A second case of elimination involves the paternal sex ratio (PSR) chromosome: a B chromosome present in some populations of arrhenotokous wasps. Only some male wasps carry this chromosome, which, when present, results in the elimination of the entire paternal genome upon fertilization, except for the PSR chromosome, and the eggs develop into male instead of female wasps. The PSR chromosomes carried by some species of *Nasonia* and *Trichogramma* genera are the most widely analyzed (Eickbush *et al.*, 1992; Van Vugt *et al.*, 2005 and references therein). The PSR from *Nasonia vitripennis* showed four satDNAs, three PSR specific and one shared with the A chromosomes. The three PSR-specific repetitive families also shared two palindromic DNA sequences that are highly conserved among the repeat families (Eickbush *et al.*, 1992). Notably, two PSR-specific repetitive families showed an open reading frame (ORF), although the transcripts have not been detected (Eickbush *et al.*, 1992). Unlike the case of *Nasonia*, in *Trichogramma kaykai* the PSR chromosomes lack chromosome-specific repeat families. The 45S rDNA seems to be the only large tandem-repetitive sequence in this case (Van Vugt *et al.*, 2005 and references therein).

Transposable elements and satellite DNA

The presence of TEs inserted in repeated DNA from *D. melanogaster* has been well known for some time (Charlesworth *et al.*, 1994). Transposition is considered one of the mechanisms of nonreciprocal transfer in the process of concerted evolution (Dover, 2002). Directly and inversely repeated sequences are very common in satDNA and they are common characteristics of many TEs. In addition, TEs or related sequences have been identified as the main component of certain satDNA families. Consequently, it has been suggested that the TEs might have contributed in some cases to the formation and spread of the satDNA. It has also been suggested that a centromeric TE or a transcript from them may participate in heterochromatin formation and in gene expression within heterochromatin (reviewed by Dimitri *et al.*, 2005).

A contribution by TEs to the origin and/or amplification/homogenization of satDNA has been suggested in several species insects, as a number of species of the *D. virilis* group (Heikkinen *et al.*, 1995). A family of TEs similar to miniature interspersed transposable elements (MITEs) has been found in the genome of *D. subobscura*

and *D. maderensis*. These elements may have produced the species-specific satDNA from the closely related species *D. guanche* (Miller *et al.*, 2000). The darkling beetle *M. goudoti* has a 1.2 kb satDNA with MITE-like sequences, suggesting that the transposition could have had an important role in its homogenization, and on its chromosomal location in all heterochromatic regions, including telomeric regions and the Y chromosome (Pons, 2004 and references therein).

Several authors (Abad and Villasante, 2000; Miller *et al.*, 2000) have suggested the conversion of TEs into a functional chromosome structure, such as telomeres or centromeres. In accordance with this suggestion, telomeric non-LTR retrotransposons have been detected in the centromeric region of the Y chromosome in different species from the *melanogaster* species group (Berloco *et al.*, 2005). *D. melanogaster* transposon BEL has been found in the functional 420-kb *Drosophila* minichromosome centromere. The transposon BEL is one of the five conserved, intact and complete transposons inserted directly into the AATAT array of this functional centromere (Sun *et al.*, 1997). The fly *C. capitata* also showed interspersed TEs in satDNA (Stratikopoulos *et al.*, 2002), some of which was highly conserved and showed a significant similarity with transposon BEL. However, it is not known whether this conservation, even in different organisms, is due to a recent insertion or it is a consequence of a selective or functional constraint probably related to centromere activity (Sun *et al.*, 1997; Stratikopoulos *et al.*, 2002).

It has been proposed that the mammalian centromeric protein B (CENP-B) is derived from a 'domesticated' *pogo*-like transposon (Casola *et al.*, 2008). In addition, CENP-B boxes, binding sites for the CENP-B protein, have a strong similarity with the terminal inverted repeats of *pogo* transposons. It has also been suggested that CENP-B protein and CENP-B boxes could have a possible role in recombining DNA sequences (reviewed in Kipling and Warburton, 1997), perhaps in sequence exchanges between satDNAs (Stitou *et al.*, 1999). The dipteran *C. pallidivittatus* and several ant species from the *Messor* genus showed conserved CENP-B boxlike motifs within of satDNA (motifs also shown by certain other satDNA insects analyzed below) and TEs, which are probably active (Rosén *et al.*, 2002; Palomeque *et al.*, 2006).

A 155-bp tandem repeat is located in all the centromeric regions of *C. pallidivittatus*. Another 375-bp tandem repeat is located exclusively in the centromere of chromosome 3 from this species. Several short interspersed element-like Cp1 elements with specific insertion sites and identical target-site duplications have been found within both centromeric repeats. Another element, the palindromic Cp80 (80-bp length) has also been found inserted into specific sites of the 155-bp repeats. Cp80 shows a sequence motif similar to the CENP-B box of mammals and a limited number of recombined forms were found, suggesting that Cp80 DNA may be a hot spot for recombination. Notably, the 155-bp repeat was also present exclusively in the telomere region of the left end of the short telocentric fourth chromosome, 4L. All of the other telomeres end in complex 340–350 bp telomere-specific repeats. A transcriptionally active ORF is located a few kilobases away from the 155-bp repeats. This ORF has degenerate inverted repeats, containing a modified

form of the Cp80 element with the putative CENP-B boxes truncated. Rosén *et al.* (2002) proposed that the 4L ORF may constitute a parallel CENP-B gene, both with an evolutionary origin in transposons. The putative product of the ORF has regions with similarities to transposase, DNA binding and endonuclease motifs (Rosén *et al.*, 2002 and references therein).

Highly conserved AT-rich 79-bp tandem repeats have been characterized from several species from the genus *Messor*. The highest sequence conservation corresponds to a region with inverted repeats that contain a CENP-B-like motif. Palomeque *et al.* (2006) reported the existence of an MITE inserted into the satDNA of *Messor bouvieri*. A *mariner*-like element was further inserted either into the satDNA within a degenerate palindrome (including the CENP-B-like motif) or into the MITE element on a specific target site. The *mariner*-like element is sometimes inserted at different positions within the satDNA but more frequently in the aforementioned position. The *mariner*-like element is transcribed and its presumed transposase is probably active. The sequence features of certain clones suggested that the *mariner*-like element might have taken part in the expansion of satDNA between chromosomes.

Evolution of satellite DNA

Concerted evolution

It is generally accepted that repeated sequences follow an evolutionary pattern known as concerted evolution. The spreading of the new variants throughout the repeats of a family leads to variant homogenization and takes place by means of a variety of genomic-turnover mechanisms, such as nonreciprocal DNA transfer within and between chromosomes (gene conversion, unequal crossing over, slippage replication, transposition, RNA-mediated exchange). The consequence of the concerted evolution is the sequence homogenization within a repeat family and their subsequent fixation in the sexual population, a process known as molecular drive (Dover, 2002 and references therein).

Different stages of transition may originate during the fixation processes of the randomly produced variants since the turnover would have to occur in a gradual manner. Strachan *et al.* (1985) reported a method to quantify these different transitional stages. Recently, this method has been applied to the study of coleopteran satDNA. Sequence data from satDNA from Iberian, Balearic and Moroccan *Pimelia* spp. suggested that the turnover mechanism has occurred in a gradual manner, according to a molecular-drive model (Pons *et al.*, 2004). However, in most satDNA, the intermediate stages of the gradual process have not been witnessed.

The importance of meiosis and chromosome segregation in the fixation of sequence variants has been supported experimentally in satDNA from the genus *Bacillus*; these organisms show very different reproductive frameworks, ranging from bisexuality to auto- and apomictic unisexuality, which allows for the uncoupling of the homogenization and fixation processes (Mantovani *et al.*, 1997) contributing to concerted evolution. Studies in this genus support the idea that sexuality can act as a driving force in the fixation of sequence variants but that the absolute values of sequence

diversity are linked to the characteristics of each species, such as copy number of the repeat and probably even number and activity of the TEs; these studies also indicate that given enough time the sequence-homogenization processes can happen in unisexual taxa (Luchetti *et al.*, 2003 and references therein).

The importance of the different reproductive strategies in the evolutionary features of satDNA has also been indicated in eusocial insects. The absence of variant fixation in satDNA from termites of the genus *Reticulitermes* could be explained by their eusocial character, since it hinders random mating and reduces the number of reproducers to a few units (Luchetti *et al.*, 2006). Eusociality and especially haplodiploidy solidly explain the relative lack of homogenization and fixation in the satDNA from ants of the genera *Formica* and *Messor*. In a haplodiploid system, the mutation rate in haploid males could counteract the effectiveness of the genome-turnover mechanism (Lorite *et al.*, 2004b and references therein).

Changes in copy number

It has been also suggested that changes in the number of copies could produce species-specific satDNA as a result of a differential amplification of preexisting repeats. Fry and Salser (1977) suggested that the ancestor of a closely related species contained a 'library' of repeat sequences, some of which could be amplified during cladogenesis. The copy number of satellite repeats could change mainly by unequal crossing-over events, although other processes, such as replication slippage, rolling-circle replication, conversion-like mechanisms and other unknown mechanisms, could also be involved (reviewed by Charlesworth *et al.*, 1994).

Mestrovic *et al.* (1998) demonstrated experimentally, for the first time to our knowledge, some of the postulates of this 'library model'. In coleopteran genera *Palorus* and *Pimelia*, it appears that a common ancestor bore the majority of the major satDNA from the present species. This satDNA would form part of 'the library of satellite sequences' (Pons *et al.*, 2004; Bruvo-Madaric *et al.*, 2007). The study of satDNA from several coleopteran species has suggested that the turnover process could have occurred at different rates. Thus, it could occur in a gradual manner, as in the satDNA from Iberian, Balearic and Moroccan *Pimelia* species, or by means of abrupt, saltatory replacement, as in satDNA from the *Pimelia* species endemic to the Canary Islands (Pons *et al.*, 2004). The gradual turnover process would imply that in some cases there should be no apparent changes for long evolutionary time, a pattern that has been described in a PRAT satDNA family, which represents the major satDNA from the coleopteran *Palorus ratzeburgii*. It is also found in low copy numbers in other species and genera, separated by a significant evolutionary period of about 50–60 Myr ago; the sequences by PCR also show high mutual similarity with ancestral mutations in all species as well as the absence of any species diagnostic mutations (Mravinac *et al.*, 2002).

The 'library model' has also been experimentally supported in insects with different reproductive frameworks; for example, the Bag320 satDNA of the bisexual *Bacillus grandii* and of the parthenogenetic *B. atticus* is

also present in bisexual and parthenogenetic *B. rossius*, although with low copy number (Cesari *et al.*, 2003). Similarly, the satDNA from the haplodiploid parasitic wasp *T. brassicae* was also detected in low copy number in several closely related species (Landais *et al.*, 2000).

Changes in copy number of satDNA associated with changes in the length of repeats and/or chromosome location have been described in several insect taxa. Of course, this evolutionary pattern cannot be solely attributed to the differential amplification of repeats. Other evolutionary processes could be involved. Evidence of an alternative process in operation comes from a satDNA family in the tropical silkworm *Antheraea mylitta*, which was detected by PCR assays in different eco races and in other silk-producing insects, although with variations in length (550–666 bp). The repetitive DNA showed an imperfect inverse repeat of 76 bp, the first repeat has a palindromic region that is absent in the second. Mahendran *et al.* (2006) suggested that the palindromic region may be a hot spot for crossing over and replication slippage, processes that probably cause the variable length of the repeat between eco races. satDNA shared by several subspecies of the dipteran *Chironomus thummi* is characterized by great differences in copy number, length of the repeats and chromosomal localization among the subspecies (Ross *et al.*, 1997).

Changes in sequence variability

The internal sequence variability of each satDNA in each species depends mainly on the ratio between mutation and homogenization/fixation rates (Dover, 2002). It has been estimated that the insect intra-specific satDNA sequence variability is 1–13% (King and Cummings, 1997). However, this value can be extremely low, as in the grasshopper *Eyprepocnemis plorans* and the parasitic wasp *T. brassicae* (about 100 and 97% sequence similarity, respectively; López-León *et al.*, 1995; Landais *et al.*, 2000), or very high, as in the *Reticulitermes* taxa of eusocial termites (68% sequence similarity; Luchetti *et al.*, 2006). In addition, different satDNA types can coexist in a species, and the sequence variability corresponding to each type can be similar or different. The two species-specific satDNA families from the grasshopper *Oxya hyla* showed similar sequence variability (Yoshimura *et al.*, 2006a). The two satDNAs from the cricket *G. bimaculatus* also showed similar sequence variability, although one is species specific and the other is present in different congeneric species (Yoshimura *et al.*, 2006b). However, the pBuM-1 and pBuM-2 evolutionarily related satDNA subfamilies from the *D. buzzatii* species cluster (*repleta* group) showed different sequence variability, indicating a slower rate of evolution of the pBuM-2 subfamily (Kuhn and Sene, 2005). This subfamily even showed high sequence conservation among geographically isolated populations from *D. gouveai*, a species included in this group (De Franco *et al.*, 2006).

Conservation of the satellite DNA sequence length

A strict conservation of satDNA-sequence length has been found within some species and between some closely related species. *Anopheles gambiae* and some related species show three 53-bp satDNA families that are highly conserved between species (Krzywinski *et al.*,

2005). Despite compartmentalization into different genomic regions (two are Y-chromosome specific and the other has a centromeric autosomal location) and remarkable sequence difference, these families display a uniformly conserved monomer length. A strict conservation of monomer length is also found in the two highly variable satDNA subfamilies from *Reticulitermes* taxa (Luchetti *et al.*, 2006). The maintenance of monomer length has been explained by the non-neutral forces of molecular drive (Dover, 2002 and references therein). It has also been suggested that the repeat length could be a critical aspect for the nucleosome positioning (or nucleosome phasing) and for the heterochromatin condensation and centromeric function (Henikoff *et al.*, 2001). However, repeats of different lengths have been found in satDNA from closely related species (Table 1 in Supplementary information). In addition, it has been reported that the insertion of several base pairs may not appreciably alter the nucleosome-phasing pattern (Simpson, 1991). Another possibility could be that the monomer-length conservation could be necessary for the modulation of higher-order structures. The human CENP-B binds to DNA as a dimer; the rigid monomer length may be needed to maintain the appropriate locations of the CENP-B boxes for protein binding (Yoda *et al.*, 1998). It is also possible that the length requirements could be a consequence of the interaction between satellite-array and specialized centromere proteins (Talbert *et al.*, 2004).

Highly conserved regions in satellite DNA

Many of the repetitive units of satDNAs have highly conserved regions, whereas other regions vary considerably. Some satellite repeats show a high degree of sequence heterogeneity but the variable sites are distributed in a nonrandom manner among the satellite monomers (Hall *et al.*, 2003). This evolutionary pattern is not only found in the satDNA of insects but also in satDNAs from other organisms such as *Arabidopsis thaliana* (Hall *et al.*, 2003). Within insect satDNA, for example, the satDNA families from the beetle genus *Tribolium* have variable and conserved segments and several common characteristics, such as short inverted repeats in the vicinity of an A–T tract, nonrandom distribution of A or T ≥ 3 tracts, and a CENP-B box-like motif, although *Tribolium* satDNA families do not share sequence similarity, monomer length or complexity (Mravinac *et al.*, 2005a). A similar conservation of certain segments has also been described in other insect satDNAs (Mravinac *et al.*, 2005b). In addition, conserved CENP-B box-like motifs have been found in other insect genera, such as *Chironomus*, *Messor* and *Formica* (Rosén *et al.*, 2002; Lorite *et al.*, 2002a, 2004b).

Probably the most studied satDNA-binding protein is the CENP-B. The CENP-B box is composed of 17 bp in human centromeric α -satellite DNA (alphoid DNA) and it contains five polymorphic sites in its consensus. In addition, the CENP-B-binding sites appear at regular intervals in human α -satellite. Ohzeki *et al.* (2002) report the loss of CENP-B-binding activity when alphoid DNA is modified by point mutations in CENP-B boxes. It has been suggested that the polymorphisms may be involved in the phasing of CENP-B boxes within the satellite, necessary for the formation of higher-order chromatin

structures (Yoda *et al.*, 1998; Choo, 2000). The implication of CENP-B–CENP-B box interaction in the centromere-assembly mechanism has been experimentally supported by Ohzeki *et al.* (2002). Proteins homologous to CENP-B have been described in many eukaryotes, and a motif similar to the CENP-B box has been found in diverse satDNAs of mammals and insects (Kipling and Warburton, 1997; Lorite *et al.*, 2004b and references therein), and CENP-B protein has been found in *D. melanogaster* (reviewed in Craig *et al.*, 1999). It is probable that CENP-B-like proteins with similar functions exist in insects. In this case, some repeat regions may be maintained by selective pressure, as other authors have suggested (Ugarkovic, 2005).

Other more variable satDNA regions may also be involved in the interaction with specific proteins. The histone H3 is replaced in centromeric nucleosomes by a special H3-like histone (CENH3). CENP-A and CID were the histone H3-like proteins found on active centromeres from human and *Drosophila*, respectively (Smith, 2002). The CENH3 protein region, which is likely to contact the centromeric DNA, appears to be under adaptive selection. Coevolution or adaptive evolution of the centromere protein has been suggested (Henikoff and Malik, 2002; Talbert *et al.*, 2004). It has been suggested that the variable regions of satDNA could also be functionally important for interaction with this type of protein, since they could shape the adaptive evolution of proteins such as CENH3. Dawe and Henikoff (2006) considered that 'the sequence variation in key kinetochore proteins is the outcome of a complex interplay between histone deposition, selfish DNA, and meiotic drive that enables the organism to maintain Mendelian segregation of the chromosomal DNA of the organism'.

Evolutionary dynamics of satellite DNA

Several hypotheses have been proposed to explain the dynamics of changes in satDNA. Henikoff and Malik (2002) suggested that the changes in repeated centromeric sequences could be due to a genetic conflict and interaction between these sequences and DNA-binding kinetochore proteins. We have already mentioned hypotheses involving the coevolution or adaptive evolution of centromere proteins (Henikoff and Malik, 2002; Talbert *et al.*, 2004) and the 'library' hypothesis (Ugarkovic and Plohl, 2002). Another theoretical model has been proposed by Nijman and Lenstra (2001), who suggested that 'the homogeneity of interacting repeat units is both cause and consequence of the rapid turnover of satDNA'. They considered a sequence of events in which, at first, the divergence of sequence variants would expand and increase. Different satDNAs with different evolutionary origins could coexist during these phases. Then the satDNA would spread over all centromeres and eventually enter a terminal phase in which the interactions between repeat units have stopped as consequence of the gradual loss of homogeneity, being replaced by a younger satDNA.

According to the literature, many satDNAs appear to have originated from short sequences present in ancestral species. During the process of species divergence, the new satellite variants are homogenized within and between chromosomes by concerted evolution. In species such as *D. virilis*, *D. simulans* and *D. melanogaster*, the repeat may have originated from a short motif (7–9 bp)

(Lohe *et al.*, 1993 and references therein). On the other hand, other insects have more complex repeats and other processes are probably involved in their origin and evolution (Charlesworth *et al.*, 1994; Dover, 2002). For example, it has been suggested that the satDNA from the parasitic wasp *T. brassicae* could have originated from an 80-bp-long unit through duplication, inversion and insertion of partially duplicated sequence elements (Landais *et al.*, 2000).

It is not known why some satDNA sequences are conserved for long evolutionary periods whereas others undergo dynamic sequence changes. As examples of conserved satDNAs, we highlight the 370-bp satDNA, which was found with sequence uniformity in eight species of the *D. virilis* group that diverged at least 20 Myr ago (Heikkinen *et al.*, 1995), and especially the dodeca satDNA from *Drosophila*, which is conserved in evolutionarily distant species such *Homo sapiens* and *Arabidopsis*. The dodeca satDNA, the 18H satellite from the centromere of *Drosophila* Y chromosome, and the evolutionarily conserved human centromeric 5-bp satellite family (the human dodeca-like satellite) form G-quartet structures (Abad and Villasante, 2000 and references therein). The *Drosophila* dodeca centromeric-binding protein contains several domains with high-affinity RNA- and ssDNA-binding motifs. It has been suggested that this protein might guide small RNAs in facilitating heterochromatin formation or, as HP1 protein, it could be involved in the maintenance of heterochromatin structure in an RNA-dependent process (reviewed by Bernstein and Allis, 2005).

Final remarks and conclusions

The major DNA components of heterochromatin of eukaryotic organism are satDNAs. The variability of satDNA sequences between species and their absence from human neocentromeres (Barry *et al.*, 1999) has raised the question of whether any specific satellite sequences are necessary to a particular centromeric function. It has been suggested that epigenetic modification governs centromere function and that the incorporation of histone variants could cause epigenetic modifications (Jin *et al.*, 2005). Nevertheless, recent data showing the presence and the maintenance of conserved and variable domains in centromere satellite sequences strongly indicate a sequence-dependent role in certain eukaryotic organisms (Hall *et al.*, 2003, 2005).

The study of insect satDNAs indicates the evolutionary conservation of certain features regardless of their sequence heterogeneity. Such features include conserved monomer length, motifs, conserved regions, and/or secondary and tertiary structures. They may act as protein-binding sites, such as structural domains or sites for epigenetic modifications. It is possible that the higher-order structures and other features may be necessary to 'expose' the centromeric chromatin outside the condensed chromosome, assuring its contact with proteins and other components necessary for centromeric function (Sun *et al.*, 2003).

It appears that even though the processes of concerted evolution have shaped insect satDNA, selective constraints are also involved. The selective constraints may be due to the satDNA sequence interaction with specific proteins important in heterochromatin formation and in

the possible role of satDNA in controlling gene expression. This evolutionary pattern is shared with other eukaryotic organisms (Hall *et al.*, 2003, 2005; Ugarkovic, 2005).

The transcription of satDNA has been described in vertebrates, invertebrates and plants. In insects, differential satDNA expression has been observed in relation to cell type, developmental stage, sex and caste of the individuals as well as in relation to other transcription differences, which would support the idea of their involvement in gene-regulation processes. In addition, the satDNA or its transcripts appear to be involved in heterochromatin formation and in chromatin-elimination processes. The transcription of satDNA has been studied in very few species. We conjecture that the transcription of satDNA in insects occurs in a high number of species, especially in those in which satDNA showed evolutionary conserved structures, although it has not yet been detected. In wild-type cells from fission yeast, certain satDNA transcripts detected through nuclear run-on assays were not found, presumably because of processing events (reviewed by Pidoux and Allshire, 2005). It is possible that in insects something similar occurs, which would make it difficult to detect the transcript.

The importance of TEs in relation to insect satDNA is shown by the presence of TEs or related sequences as a constituent of satDNA in several species of insects. In addition, they may be involved in the formation of centromeres and telomeres and in the homogenization and expansion of satDNA. The existence of probably active TEs inserted into insect repetitive DNA has also been reported.

The literature on insect satDNAs shows that the knowledge of this type of DNA in insects is still fragmented and insufficient. That is, satDNAs have been studied in only a few of nearly a million insect species. With few exceptions, the satDNA has been studied in only a few species within of each taxonomic group. Since the life cycle and reproductive strategies are extremely varied in insects, the comparison between satDNAs from different groups could be particularly helpful in understanding the function and evolution of these repeated sequences.

Acknowledgements

The Spanish CGL2006-04331 project and the PAI CECJA CVI 220 supported this work. We are grateful for the valuable comments and suggestions on the paper given by three anonymous referees.

References

- Abad JP, Agudo M, Molina I, Losada A, Ripol P, Villasante A (2000). Pericentromeric regions containing 1.688 satellite DNA sequences show anti-kinetochore antibody staining in prometaphase chromosomes of *Drosophila melanogaster*. *Mol Gen Genet* **264**: 371–377.
- Abad JP, Villasante A (2000). Searching for a common centromeric structural motif: *Drosophila* centromeric satellite DNAs show propensity to form telomeric-like unusual DNA structures. *Genetica* **109**: 71–75.
- Aravin AA, Lagos-Quintana M, Yalcin A, Zabolan M, Marks D, Gaasterland T *et al.* (2003). The small RNA profile during *Drosophila melanogaster* development. *Dev Cell* **5**: 337–350.
- Bachmann L, Sperlich D (1993). Gradual evolution of a specific satellite DNA family in *Drosophila ambigua*, *D. tristis*, and *D. obscura*. *Mol Biol Evol* **10**: 647–659.
- Barceló F, Gutierrez F, Barjau I, Portugal J (1998). A theoretical perusal of the satellite DNA curvature in tenebrionid beetles. *J Biomol Struct Dyn* **16**: 41–50.
- Barceló F, Pons J, Petitpierre E, Barjau I, Portugal J (1997). Polymorphic curvature of satellite DNA in three subspecies of the beetle *Pimelia sparsa*. *Eur J Biochem* **244**: 318–324.
- Barry AE, Howman EV, Cancilla MR, Saffery R, Choo KH (1999). Sequence analysis of an 80 kb human neocentromere. *Human Mol Genet* **8**: 217–227.
- Berlaco M, Fanti L, Sheen F, Levis RW, Pimpinelli S (2005). Heterochromatic distribution of HeT-A- and TART-like sequences in several *Drosophila* species. *Cytogenet Genome Res* **110**: 124–133.
- Bernstein E, Allis CD (2005). RNA meets chromatin. *Genes Dev* **19**: 1635–1655.
- Beye M, Hasselmann M, Fondrk MK, Page RE, Omholt SW (2003). The gene *csd* is the primary signal for sexual development in the honeybee and encodes an SR-type protein. *Cell* **114**: 419–429.
- Bizzaro D, Manicardi GC, Bianchi U (1996). Chromosomal localization of a highly repeated *EcoRI* DNA fragment in *Megoura viciae* (Homoptera, Aphididae) by nick translation and fluorescence *in situ* hybridization. *Chromosome Res* **4**: 392–396.
- Blanchetot A (1991). Genetic variability of a satellite sequence in the dipteran *Musca domestica*. *EXS* **58**: 106–112.
- Bonaccorsi S, Lohe A (1991). Fine mapping of satellite DNA sequences along the Y chromosome of *Drosophila melanogaster*: relationships between satellite sequences and fertility factors. *Genetics* **129**: 177–189.
- Bruvo-Madaric B, Plohl M, Ugarkovic D (2007). Wide distribution of related satellite DNA families within the genus *Pimelia* (Tenebrionidae). *Genetica* **130**: 35–42.
- Carrera P, Azorin F (1994). Structural characterization of intrinsically curved AT-rich DNA sequences. *Nucleic Acids Res* **22**: 3671–3680.
- Casola C, Hucks D, Feschotte C (2008). Convergent domestication of *pogo*-like transposases into centromere-binding proteins in fission yeast and mammals. *Mol Biol Evol* **25**: 29–41.
- 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, Sniegowsky P, Stephan W (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**: 215–220.
- Choo KH (2000). Centromerization. *Trends Cell Biol* **10**: 182–188.
- Craig JM, Earnshaw WC, Vagnarelli P (1999). Mammalian centromeres: DNA sequence, protein composition, and role in cell cycle progression. *Exp Cell Res* **246**: 249–262.
- Dawe RK, Henikoff S (2006). Centromeres put epigenetics in the driver's seat. *Trends Biochem Sci* **31**: 662–669.
- De Franco FF, Kuhn GC, de Sene FM, Manfrin MH (2006). Conservation of pBuM-2 satellite DNA sequences among geographically isolated *Drosophila gouveai* populations from Brazil. *Genetica* **128**: 287–295.
- Dimitri P, Corradini N, Rossi F, Verni F (2005). The paradox of functional heterochromatin. *Bioessays* **27**: 29–41.
- Dimitrov S, Wolfe AP (1996). Remodeling somatic nuclei in *Xenopus laevis* egg extracts: molecular mechanisms for the selective release of histones H1 and H1 (0) from chromatin and the acquisition of transcriptional competence. *EMBO J* **15**: 5897–5906.
- Dover G (2002). Molecular drive. *Trends Genet* **18**: 587–589.

- Eickbush DG, Eickbush TH, Werren JH (1992). Molecular characterization of repetitive DNA sequences from a B chromosome. *Chromosoma* **101**: 575–583.
- 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.
- Hall SE, Kettler G, Preuss L (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.
- Heikkinen E, Launonen V, Muller E, Bachmann L (1995). The pVB370 BamHI satellite DNA family of the *Drosophila virilis* group and its evolutionary relation to mobile dispersed genetic pDv elements. *J Mol Evol* **41**: 604–614.
- Henikoff S, Ahmad K, Malik HS (2001). The centromere paradox, stable inheritance with rapidly evolving DNA. *Science* **293**: 1098–1102.
- Henikoff S, Malik HS (2002). Centromeres: selfish drivers. *Nature* **417**: 227.
- Hoskins RA, Carlson JW, Kennedy C, Acevedo D, Evans-Holm M, Frise E *et al.* (2007). Sequence finishing and mapping of *Drosophila melanogaster* heterochromatin. *Science* **15**: 1625–1628.
- Jin J, Cai Y, Li B, Conaway RC, Workman JL, Conaway JW *et al.* (2005). In and out: histone variant exchange in chromatin. *Trends Biochem Sci* **30**: 680–687.
- King LM, Cummings MP (1997). Satellite DNA repeat sequence variation is low in three species of burying beetles in the genus *Nicrophorus* (Coleoptera: Silphidae). *Mol Biol Evol* **14**: 1088–1095.
- Kipling D, Warburton PB (1997). Centromeres, CENP-B and Tigger too. *Trends Genet* **13**: 141–144.
- Krzywinski J, Sangare D, Besansky NJ (2005). Satellite DNA from the Y chromosome of the malaria vector *Anopheles gambiae*. *Genetics* **169**: 185–196.
- Kuhn GC, Sene FM (2005). Evolutionary turnover of two pBuM satellite DNA subfamilies in the *Drosophila buzzatii* species cluster (*repleta* group): from alpha to alpha/beta arrays. *Gene* **349**: 77–85.
- Kurek R, Reugels AM, Lammermann U, Bünemann H (2000). Molecular aspects of intron evolution in dynein encoding mega-genes on the heterochromatic Y chromosome of *Drosophila* sp. *Genetica* **109**: 113–123.
- Landais I, Chavigny P, Castagnone C, Pizzol J, Abad P, Vanlerberghe-Masutti F (2000). Characterization of a highly conserved satellite DNA from the parasitoid wasp *Trichogramma brassicae*. *Gene* **255**: 65–73.
- Lobov IB, Tsutsui K, Mitchell AR, Podgornaya OI (2001). Specificity of SAF-A and lamin B binding *in vitro* correlates with the satellite DNA binding state. *J Cell Biochem* **83**: 218–229.
- Lohe AR, Hilliker AJ, Roberts PA (1993). Mapping simple repeated DNA sequences in heterochromatin of *Drosophila melanogaster*. *Genetics* **134**: 1149–1174.
- López-León MD, Vazquez P, Hewitt GM, Camacho JP (1995). Cloning and sequence analysis of an extremely homogeneous tandemly repeated DNA in the grasshopper *Eyprepocnemis plorans*. *Heredity* **75**: 370–375.
- Lorite P, Carrillo JA, Aguilar JA, Palomeque T (2004a). Isolation and characterization of two families of satellite DNA with repetitive units of 135 bp and 2.5 kb in the ant *Monomorium subopacum* (Hymenoptera, Formicidae). *Cytogenet Genome Res* **105**: 83–92.
- Lorite P, Carrillo JA, Tinaut A, Palomeque T (2002a). Comparative study of satellite DNA in ants of the *Messor* genus (Hymenoptera, Formicidae). *Gene* **297**: 113–122.
- Lorite P, Carrillo JA, Tinaut A, Palomeque T (2004b). Evolutionary dynamics of satellite DNA in species of the genus *Formica* (Hymenoptera, Formicidae). *Gene* **332**: 159–168.
- Lorite P, Palomeque T, Garnería I, Petitpierre E (2001). Characterization and chromosome location of satellite DNA in the leaf beetle *Chrysolina americana* (Coleoptera, Chrysomelidae). *Genetica* **110**: 143–150.
- Lorite P, Renault S, Rouleux-Bonnin F, Bigot S, Periquet G, Palomeque T (2002b). Genomic organization and transcription of satellite DNA in the ant *Aphaenogaster subterranea* (Hymenoptera, Formicidae). *Genome* **45**: 609–616.
- 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.
- Mahendran B, Acharya C, Dash R, Ghosh SK, Kundu SC (2006). Repetitive DNA in tropical tasar silkworm *Antheraea mylitta*. *Gene* **29**: 51–57.
- Mandrioli M, Bizzaro D, Manicardi GC, Gionghi D, Bassoli L, Bianchi U (1999). Cytogenetic and molecular characterization of a highly repeated DNA sequence in the peach potato aphid *Myzus persicae*. *Chromosoma* **108**: 436–442.
- Mantovani B, Tinti F, Bachmann L, Scali V (1997). The Bag320 satellite DNA family in *Bacillus* stick insects (Phasmatodea): different rates of molecular evolution of highly repetitive DNA in bisexual and parthenogenic taxa. *Mol Biol Evol* **14**: 1197–1205.
- Matsugami A, Tani K, Ouhashi K, Uesugi S, Morita M, Ohyama T *et al.* (2006). Structural property of DNA that migrates faster in gel electrophoresis, as deduced by CD spectroscopy. *Nucleosides Nucleotides Nucleic Acids* **25**: 417–425.
- Mestrovic N, Plohl M, Mravinac B, Ugarkovic D (1998). Evolution of satellite DNAs from the genus *Palorus*—experimental evidence for the ‘library’ hypothesis. *Mol Biol Evol* **15**: 1062–1068.
- 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.
- Mravinac B, Plohl M, Mestrovic N, Ugarkovic D (2002). Sequence of PRAT satellite DNA ‘frozen’ in some coleopteran species. *J Mol Evol* **54**: 774–783.
- Mravinac B, Plohl M, Ugarkovic D (2005b). Preservation and high sequence conservation of satellite DNAs suggest functional constraints. *J Mol Evol* **61**: 542–550.
- Mravinac B, Ugarkovic E, Franjevic D, Plohl M (2005a). Long inversely oriented subunits form a complex monomer of *Tribolium brevicornis* satellite DNA. *J Mol Evol* **60**: 513–525.
- Ner SS, Blank T, Perez-Paralle ML, Grigliatti TA, Becker PB, Travers AA (2001). HMG-D and histone H1 interplay during chromatin assembly and early embryogenesis. *J Biol Chem* **276**: 37569–37576.
- Nijman IJ, Lenstra JA (2001). Mutation and recombination in cattle satellite DNA: a feedback model for the evolution of satellite DNA repeats. *J Mol Evol* **52**: 361–371.
- Ohzeki JI, Nakano M, Okada T, Masumoto H (2002). CENP-B box is required for the novo centromere chromatin assembly on human alphoid DNA. *J Cell Biol* **159**: 765–775.
- Palomeque T, Carrillo JA, Muñoz-López M, Lorite P (2006). Detection of a *mariner*-like element and a miniature inverted-repeat transposable element (MITE) associated with the heterochromatin from ants of the genus *Messor* and their possible involvement for satellite DNA evolution. *Gene* **371**: 194–205.
- Palomeque T, Muñoz-López M, Carrillo JA, Lorite P (2005). Characterization and evolutionary dynamics of a complex family of satellite DNA in the leaf beetle *Chrysolina carnifex* (Coleoptera, Chrysomelidae). *Chromosome Res* **13**: 795–807.
- Pidoux AL, Allshire RC (2005). The role of heterochromatin in centromere function. *Philos Trans R Soc Lond B Biol Sci* **360**: 569–579.
- Piergentili R, Bonaccorsi S, Raffa DG, Pisano C, Hackstein JHP, Mencarelli C (2004). Autosomal control of the Y-chromosome kl-3 loop of *Drosophila melanogaster*. *Chromosoma* **113**: 188–196.

- Plohl M, Mestrovic N, Bruvo B, Ugarkovic D (1998). Similarity of structural features and evolution of satellite DNAs from *Palorus subdepressus* (Coleoptera) and related species. *J Mol Evol* **46**: 234–239.
- Pons J (2004). Cloning and characterization of a transposable-like repeat in the heterochromatin of the darkling beetle *Misolampus goudoti*. *Genome* **47**: 769–774.
- Pons J, Bruvo B, Petitpierre E, Plohl M, Ugarkovic D, 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.
- Pons J, Bucur R, Vogler AP (2003). Higher-order repeats in the satellite DNA of the cave beetle *Pholeuon proserpinae glaciale* (Coleoptera: Cholevidae). *Heredity* **139**: 28–34.
- Rojas AA, Vázquez-Tello A, Ferbeyre G, Venanzetti F, Bachmann L, Paquin B *et al.* (2000). Hammerhead-mediated processing of satellite pDo500 family transcripts from *Dolichopoda* cave crickets. *Nucleic Acids Res* **28**: 4037–4043.
- Rojas-Rousse D, Bigot Y, Periquet G (1993). DNA insertions as a component of the evolution of unique satellite DNA families in two genera of parasitoid wasps: *Diadromus* and *Eupelmus* (Hymenoptera). *Mol Biol Evol* **10**: 383–396.
- Rosén M, Castillejo-López C, Edström JE (2002). Telomere terminating with centromere-specific repeats is closely associated with a transposon derived gene in *Chironomus pallidivittatus*. *Chromosoma* **110**: 532–541.
- Ross R, Hankeln T, Schmidt ER (1997). Complex evolution of tandem-repetitive DNA in the *Chironomus thummi* species group. *J Mol Evol* **44**: 321–326.
- Rouleux-Bonnin F, Bigot S, Bigot Y (2004). Structural and transcriptional features of *Bombus terrestris* satellite DNA and their potential involvement in the differentiation process. *Genome* **47**: 877–888.
- Schittenhelm RB, Heeger S, Althoff F, Walter A, Heidmann S, Mechler K *et al.* (2007). Spatial organization of a ubiquitous eukaryotic kinetochore protein network in *Drosophila* chromosomes. *Chromosoma* **116**: 385–402.
- Simpson RT (1991). Nucleosome positioning: occurrence, mechanisms, and functional consequences. *Prog Nucleic Acid Res Mol Biol* **40**: 143–184.
- Smith CD, Shu S, Mungall CJ, Karpen GH (2007). The release 5.1 annotation of *Drosophila melanogaster* heterochromatin. *Science* **15**: 1586–1591.
- Smith MM (2002). Centromeres and variant histones: what, where, when and why? *Curr Opin Cell Biol* **14**: 279–285.
- Staiber W (2002). Isolation of a new germ line-specific repetitive DNA family in *Acricotopus* by microdissection of polytenized germ line-limited chromosome sections from a permanent larval salivary gland preparation. *Cytogenet Genome Res* **98**: 210–215.
- Staiber W (2004). Molecular evolution of homologous gene sequences in germ line-limited and somatic chromosomes of *Acricotopus*. *Genome* **47**: 732–741.
- Staiber W (2006). Chromosome elimination in germ line-soma differentiation of *Acricotopus lucidus* (Diptera, Chironomidae). *Genome* **49**: 269–274.
- Stitou S, Díaz de la Guardia R, Jiménez R, Burgos M (1999). Isolation of a species-specific satellite DNA with a novel CENP-B-like box from the North African rodent *Lemniscomys barbarus*. *Exp Cell Res* **250**: 381–386.
- Strachan T, Webb D, Dover GA (1985). Transition stages of molecular drive in multiple-copy DNA families in *Drosophila*. *EMBO J* **4**: 1701–1708.
- Stratikopoulos EE, Augustinos AA, Gariou-Papalexioi A, Zacharopoulou A, Mathiopoulou KD (2002). Identification and partial characterization of a new *Ceratitiss capitata*-specific 44-bp pericentromeric repeat. *Chromosome Res* **10**: 287–295.
- Sun X, Le HD, Wahlstrom JM, Karpen GH (2003). Sequence analysis of a functional *Drosophila* centromere. *Genome Res* **13**: 182–194.
- Sun X, Wahlstrom JM, Karpen GH (1997). Molecular structure of a functional *Drosophila* centromere. *Cell* **9**: 1007–1019.
- Talbert PB, Bryson TD, Henikoff S (2004). Adaptive evolution of centromere proteins in plants and animals. *J Biol* **3**: 18.
- Trapitz P, Wlaschek M, Bunemann H (1988). Structure and function of Y chromosomal DNA. II. Analysis of lampbrush loop associated transcripts in nuclei of primary spermatocytes of *Drosophila hydei* by *in situ* hybridization using asymmetric RNA probes of four different families of repetitive DNA. *Chromosoma* **96**: 159–170.
- Ugarkovic D (2005). Functional elements residing within satellite DNAs. *EMBO Rep* **6**: 1035–1039.
- Ugarkovic D, Plohl M (2002). Variation in satellite DNA profiles—causes and effects. *EMBO J* **21**: 5955–5959.
- Usakin L, Abad J, Vagin VV, De Pablos B, Villasante A, Gvozdev VA (2007). Transcription of the 1.688 satellite DNA family is under the control of RNA interference machinery in *Drosophila melanogaster* ovaries. *Genetics* **176**: 1343–1349.
- Van Vugt JJFA, de Nooijer S, Stouthamer R, de Jong H (2005). NOR activity and repeat sequences of the paternal sex ratio chromosome of the parasitoid wasp *Trichogramma kaykai*. *Chromosoma* **114**: 410–419.
- Willard HF, Wayne JS (1987). Hierarchical order in chromosome-specific human alpha satellite DNA. *Trends Genet* **3**: 192–198.
- Yoda K, Ando S, Okuda A, Kikuchi A, Okazaki T (1998). *In vitro* assembly of the CENP-B/alpha-satellite DNA/core histone complex: CENP-B causes nucleosome positioning. *Genes Cells* **3**: 533–548.
- Yoshimura A, Nakata A, Kuro-o M, Obara Y, Ando Y (2006a). Molecular cytogenetic characterization and chromosomal distribution of the satellite DNA in the genome of *Oxya hyla intricata* (Orthoptera: Catantopidae). *Cytogenet Genome Res* **112**: 160–165.
- Yoshimura A, Nakata A, Mito T, Noji S (2006b). The characteristics of karyotype and telomeric satellite DNA sequences in the cricket, *Gryllus bimaculatus* (Orthoptera, Gryllidae). *Cytogenet Genome Res* **112**: 329–336.

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