

Intra-specific variability and unusual organization of the repetitive units in a satellite DNA from *Rana dalmatina*: Molecular evidence of a new mechanism of DNA repair acting on satellite DNA

Isidoro Feliciello^a, Orfeo Picariello^b, Gianni Chinali^{a,*}

^a Dipartimento di Medicina Clinica e Sperimentale, Università di Napoli Federico II, Via Pansini 5, I-80131 Napoli, Italy

^b Dipartimento di Biologia Strutturale e Funzionale, Università di Napoli Federico II, Monte S. Angelo, Via Cinthia, I-80126 Napoli, Italy

Received 7 April 2006; received in revised form 6 July 2006; accepted 13 July 2006

Available online 28 July 2006

Received by M. Di Giulio

Abstract

We have characterized the S1 satellite from eight European populations of *Rana dalmatina* by Southern blot, cloning and a new method that determines the sequence variability of repetitive units in the genome. This report completes our previous studies on this satellite DNA family, thus providing the first characterization of the overall variability of the structure and genomic organization of a satellite DNA within a species and among related species.

The S1 satellite from *R. dalmatina* has a pericentromeric location on ten chromosome pairs and presents two homologous repeats S1a (494 bp) and S1b (332 bp), mostly organized as composite S1a–S1b repetitive units. In other brown frog species, both repeats have different sequences and locations, and are usually organized as separate arrays, although composite S1a–S1b repeats represent a minor, widely variable component in *Rana italica*. The average genomic sequences indicate that the species contains an enormous number of variants of each repeat derived from a unique, species-specific common sequence. The repeat variability is restricted to specific base changes in specific sequence positions in all population samples.

Our data show that the structure and evolution of S1 satellite family is not due to crossing-over and gene conversion, but to a mechanism that maintains the ability of the satellite DNA to assemble in constitutive heterochromatin by replacing altered satellite segments with new arrays generated by rolling circle amplification. The mode of action of this repair process not only directly explains the intra- and inter-specific variability of the structure and organization of the S1 satellite repeats from European brown frogs, but also accounts for all general features of satellite DNA in eukaryotes, including its discontinuous evolution. This repair mechanism can maintain the satellite structure in a species indefinitely, but also promote a rapid generation of new variants or types of satellite DNA when environmental conditions favor the formation of new species.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Satellite DNA; Rolling circle; Homologous recombination; DNA repair; Heterochromatin; Evolution

1. Introduction

Satellite DNA is a consistent part of the genome of most eukaryotes and is characterized by a large number of repeated sequences organized in long tandem arrays. It is mostly located

Abbreviations: bio-16-dUTP; biotin-16-2'-deoxy-uridine-5'-triphosphate; bp; base pair(s); kb; kilobase(s) or 1000 bp; PCR; polymerase chain reaction; MCS; most common sequence; *R.*; *Rana*; sat-proteins; satellite DNA-binding proteins.

* Corresponding author. Tel.: +39 081 7463355.

E-mail address: chinali@unina.it (G. Chinali).

in the regions of constitutive heterochromatin located in or near centromeres and telomeres. The repetitive units of satellite DNA vary greatly in size and sequence, similar satellite DNAs being found only in closely related species (John and Miklos, 1979; Brutlag, 1980). No function has been definitely assigned to this highly repetitive DNA, although it has been implicated in the genomic structure and evolutionary processes (Bostock, 1980). Centromeric satellite DNA has also been implicated in centromere function (Henikoff, 2000).

The mechanisms underlying the origin, evolution and maintenance of satellite DNA are poorly understood. Many

theoretical models have been proposed (e.g. Smith, 1976; Dover, 1982; Stephan, 1989; Modi, 1993; Charlesworth et al., 1994), but all lack substantial experimental support. Indeed, the arrangement of nearly identical repeats in long tandem arrays, hampers the study of any event of recombination, amplification, mutation and substitution that may occur in this repetitive DNA.

Differently, the S1 satellite DNA of the European brown frogs, which was first characterized in *Rana italica* (Cardone et al., 1997), lends itself to the study of most of these features. Thanks to the presence of two homologous repetitive units, S1a (494 bp) and S1b (285 bp), in the same satellite DNA, we were able to identify extensive differences in the organization of these repeats in the genome of frogs from the same population. This provided the first direct experimental evidence of specific processes that determine rapid changes in the organization of the repetitive units in a satellite DNA. More recently, we found that the mixture of the S1a or S1b repetitive units amplified by PCR from the genomic DNA of *Rana graeca* could be directly sequenced to yield an average sequence of these two repeats that is representative of all the repetitive units present in the genome (Picariello et al., 2002). This procedure has been exploited to obtain a detailed characterization of the intra-specific variability of the S1a repeat from *Rana temporaria* (Feliciello et al., 2005).

Herein, we describe a series of features identified in the S1 satellite DNA of *Rana dalmatina*. This report confirms, extends and completes our previous studies on this repetitive DNA. The combined analysis of the previously known and newly determined characteristics of satellite DNA clearly indicates the presence of a mechanism of DNA repair that acts specifically on this repetitive DNA. Unlike all other known repair systems, this new mechanism does not directly repair mutations in DNA; rather it restores the functionality of satellite DNA by replacing damaged segments with newly synthesized segments. This replacement-repair mechanism can account not only for the maintenance of a definite structure and organization of the satellite repetitive units within each species, but also for the rapid fixation of specific variants of the repetitive DNA in new species.

2. Materials and methods

2.1. Animals

We analyzed tissue specimens of *R. dalmatina* from populations A–H of the following areas: Girifalco, Cosenza province, Calabria, Italy (A); Teggiano, Salerno province, Campania, Italy (B); San Felice, Sondrio province, Lombardia, Italy (C); Wachtberg, NRW, Germany (D); Karlskrona, Sweden (E); Køge, Copenhagen, Denmark (F); Tuhan, Melnik district, Czech Republic (G); Parga, Igoumenitsa province, Epyrus, Greece (H).

2.2. DNA isolation

Highly purified DNA was extracted from blood, skin or liver, as previously described (Cardone et al., 1997). Results were independent of the tissue used for DNA preparation. Tissue samples from population D (animals found dead) and population E (animals kept in ethanol for about 40 years) contained

highly degraded DNA. The larger DNA fragments present in these samples (about 0.6–1.2 kb in size) were isolated by gel-filtration on Sepharose 4B (Pharmacia, Uppsala, Sweden) and used for PCR amplification.

DNA from clones of S1 satellite DNA in pTZ19R plasmid vector (Pharmacia, Uppsala, Sweden) was isolated by a modified alkaline lysis method (Feliciello and Chinali, 1993).

2.3. Molecular cloning

DNA fragments containing whole S1a or S1b repeats were obtained by digestion of genomic DNA of *R. dalmatina* from population B with *EcoRV*. These fragments were cloned in the *SmaI* site of pTZ19R as described elsewhere (Cardone et al., 1997). Fragments produced by digestion with *Asp718I* (*KpnI*) and containing whole or partial S1a or S1a–b repeats were cloned in the same site of pTZ19R.

2.4. PCR amplifications

Because of the tandem organization of satellite DNA, whole S1a and S1b repetitive units from *R. dalmatina* could be amplified by PCR using primers with the same origin in the repeat sequence, but with opposite orientation. Four sets of primer pairs were used. The two primer sets Rdr371+Rin372 and Rdr160+Rin161 are described elsewhere (Picariello et al., 2002). Two other primer pairs were also used: Rdr05 (AGGTCCTARCACCAAAGG)+Rin07 (CCTCCAAAC TACCCACAT) and Rdr273 (CCACATTGTAGCCCCATA)+Rin274 (GCRAARAAAGYAAGTCGC) located in positions 112–144 and 378–412 of the S1a repeat, respectively (see Fig. 2). Procedures for amplification of genomic DNA with a primer pair, and isolation by preparative gel electrophoresis of the amplification products corresponding to the S1a or S1b monomer units are reported elsewhere (Picariello et al., 2002). The primer pairs Rdr05+Rin07 and Rdr160+Rin161 amplified the S1a repeat only, while the primer pairs Rdr273+Rin274 and Rdr371+Rin372 amplified predominantly the S1b repeat.

2.5. DNA sequencing

After purification, the PCR-amplified S1a and S1b repeats were sequenced in both orientations by an automatic sequencer (Applied Biosystems, Foster City, Calif.) using the BigDye Terminators kit by the same manufacturer and the primers used for amplification. Each type of repeat was amplified and sequenced with two different primer pairs. In this way, the portion of the S1a or S1b sequence around each primer pair was determined by sequencing the same repeat amplified by the other primer pair.

Each amplification product originates from a very large number of similar, but not identical S1a or S1b unit repeats. As expected, sequence electropherograms revealed more than one base in several positions of the sequence. The percentage of a base in a position of the sequence was assumed to correspond to the average of its percentages of total signals measured in this position in the direct and inverse sequence electropherograms. Only base pairs found in both sequencing orientations were

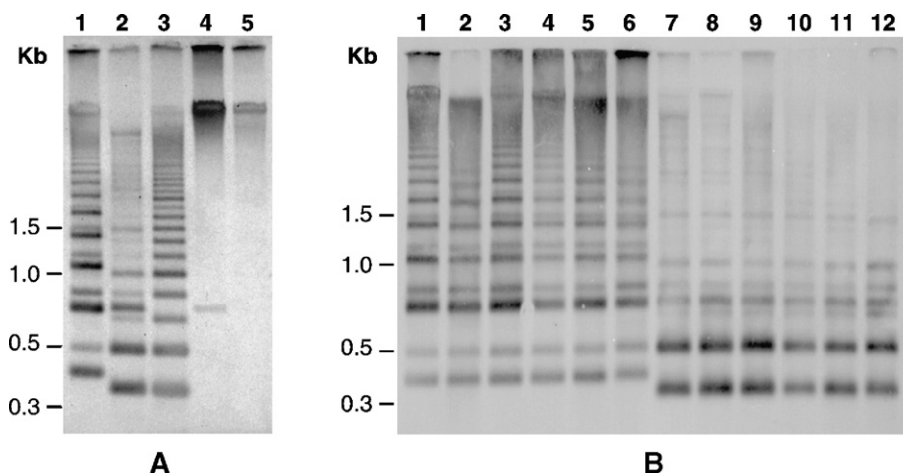


Fig. 1. Southern blot of the genomic DNA from *R. dalmatina* digested with various restriction enzymes. A: genomic DNA (3 μ g) from a specimen of *R. dalmatina* from population H (Greece) was digested with *KpnI* (lane 1), *EcoRV* (lane 2), *NdeI* (lane 3), *NheI* (lane 4), and *StuI* (lane 5). After electrophoresis on 1.5% agarose gel (1.5 h at 4 V/cm), samples were analyzed by Southern blot hybridization with a biotin-labelled probe from the S1a satellite DNA of *R. italica*. B: genomic DNA (3 μ g) of specimens from populations A, B, C, F, G and H of *R. dalmatina* was digested with *KpnI* (lanes 1 to 6) or *EcoRV* (lanes 7 to 12) and analyzed by Southern blot hybridization as in panel A.

considered. The sequence resulting from the analysis of two amplification products of the same repetitive unit amplified by two different primer pairs was defined “average genomic sequence” of the repeat.

For submission to the GenBank, the average genomic sequence was converted to a “genomic consensus sequence” using a single letter code (e.g.: Y for C or T; B for C, G or T) to identify multiple bases in the same position. Threshold limits were imposed for inclusion of a minor base in the consensus sequence. A minor base was included only if it gave a signal at least one-tenth of that of the main base, both in the direct and inverse sequencing orientation. We adopted threshold limits lower than those previously adopted for *R. graeca* (Picariello et al., 2002) because of the better chemistry of reaction (BigDye vs. Dye terminators).

The S1a, S1b and S1a–S1b units cloned in pTZ19R were sequenced using pUC18 primers.

2.6. Biotin labelled probes

Biotinylated probes of the S1a or S1b repetitive unit were obtained by PCR using bio-16-dUTP (Sigma, St. Louis, Mo.) and purified S1a or S1b repeats of *R. dalmatina*, or S1a repeat from *R. italica*, amplified from genomic DNA, as templates (Cardone et al., 1997). The S1a probes from the two frog species are highly homologous and produced equivalent results.

2.7. Southern blot and dot blot analyses

Southern blot and quantitative dot blot analyses were carried out as described (Picariello et al., 2002) using S1a biotinylated probes.

2.8. Cytological and in situ hybridization procedures

FISH experiments were carried out as previously described (Cardone et al., 1997) using blood cells from *R. dalmatina* and a biotinylated probe of S1a repeat from the same species.

3. Results

3.1. Southern blot analysis

We have firstly characterized the S1 satellite DNA from *R. dalmatina* by Southern blot analysis of genomic DNA digested with *KpnI*, *EcoRV*, *NdeI*, *StuI*, or *NheI*, the five restriction enzymes found to be present in the same repetitive DNA in other brown frog species. As illustrated in Fig. 1A, the S1 satellite DNA of this species does not contain sites for *NheI* (lane 4) and *StuI* (lane 5). The other three restriction enzymes produced a complex pattern of hybrid bands very unusual for a satellite DNA. The faster migrating bands could be easily interpreted on the basis of our previous studies. The 0.49-kb band in lanes 2 and 3 and the 0.39–0.49-kb doublet in lane 1 indicates the presence of an S1a repeat containing one site for *EcoRV* and *NdeI*, and two sites for *KpnI*, respectively. Moreover, the 0.33-kb band (putative monomer) and the 0.66-kb band (putative dimer) in lanes 2 and 3, indicate presence of a species-specific S1b repeat of about 0.33 kb containing one *EcoRV* and one *NdeI* site.

The origin of some other bands present on Southern blots is not immediately evident. In addition to the 0.39–0.49-kb S1a doublet, a series of band doublets is present in lane 1 spaced at regular intervals that mimic the typical ladder of a satellite DNA. Interestingly, each doublet is 0.33 kb larger than the preceding one, 0.33 kb being the size of the S1b repeat. If one considers that this repeat contains no *KpnI* sites, the pattern indicates that most S1a and S1b repeats are organized as composite repetitive units in which the two repeats are present in various combinations. However, the presence of an undigested fraction of S1 satellite DNA in lane 1 and of an S1a ladder extending up to a 4-mer in lane 2 indicates that a minor fraction of the S1b and S1a repeats is organized as separate tandem arrays. This organization is opposite of that found in *R. italica* in which the organization as separate arrays predominates (Cardone et al., 1997). Only the faint hybrid band

of about 0.73 kb in lane 4, and the more pronounced band of the same size in lanes 1 and 2 are difficult to assign. We have attempted to clone this fragment from genomic DNA digested with *KpnI* or *EcoRV* without success. These bands likely originate from a dispersed repetitive element that has sequence homology with the S1 satellite DNA and contain sites for *KpnI*, *EcoRV* and *NheI*, as suggested by the fact that no oligomer of this band has been observed under conditions of partial digestion with any of these restriction enzymes.

The same pattern of satellite DNA hybrid bands was observed in all the population samples tested. This is illustrated in Fig. 1B for the genomic DNAs of six population samples digested with *KpnI* or *EcoRV*. This experiment does not include the samples of the two populations from Germany and Sweden because the DNA was too degraded. After isolation by gel filtration, the digestion of the larger DNA fragments (0.6–1.1 kb) present in these two samples produced the same hybrid bands observed with not degraded DNA up to 1 kb, the testable size range (results not illustrated). This confirmed that the structure and organization of the repetitive units are constant within the species.

3.2. Variability of S1a and S1b repeats in the species

We used two methods to sequence the S1a and S1b repetitive motifs of *R. dalmatina*. One is the conventional cloning procedure that yields the sequence of individual repetitive units (see following section of Results). The other method, recently

described by us, uses PCR amplification of S1a and/or S1b repeats from genomic DNA by pairs of primers with an opposite orientation and different origin. This produces a mixture of amplified repeats that is representative of all repeats present in a genome (Picariello et al., 2002). Direct automated sequencing of the resulting mixture of purified S1a (or S1b) monomers yields an “average genomic sequence” that represents the overall variability of the repeat in the genome tested.

We have determined the genomic sequences of the S1a and S1b repeats amplified from DNA samples from eight different European population of *R. dalmatina*. In all these sequences, more than 80% of the positions show a unique base, while the other positions contain a main base and usually only one minor base in varying percentages. These sequences can be analyzed in various ways. The simplest analysis considers only the main base present in each position of sequence. This defines the “Most Common Sequence” (MCS) of the repeat in the genome tested. A striking result is obtained: all the population samples tested yield an identical MCS both for the S1a and the S1b repeat.

The MCS of the S1a and S1b repeat of *R. dalmatina* are illustrated and compared in Fig. 2. Unlike the S1b repeats of *R. italica* and *R. graeca*, the S1b repeat of *R. dalmatina* is not homologous to the S1a repeat throughout its entire length. Most of the non-homologous part duplicates an immediately adjacent sequence, thereby producing a 30-bp long palindrome. As already observed in *R. italica* and *R. graeca*, the sequence of the S1b repeat is not identical to that of the homologous part of



Fig. 2. Alignment of the most common sequences (MCS) of the S1a and the S1b repeat of *R. dalmatina*. Arrows indicate the 30-bp inverted repeat present in S1b. In positions presenting a different base, an analysis of the average genomic sequences was carried out to verify the eventual presence in one repeat of a minor base corresponding to the major or minor base present in the same position in the other repeat. Lower case and capital letters indicate positions in which a common base is present or absent, respectively. The sites for the restriction enzymes *KpnI*, *EcoRV* and *NdeI* are underlined.

the S1a repeat (bp 243–494 and 1–44), and presents 19 base changes. As indicated by capital letters, eight of these changes are complete, i.e. in these positions the sequence electropherograms of the two repeats present no common base at all. The relevance of this feature is analyzed in the Discussion section.

The MCS of S1a and S1b repeat contains all the restriction sites indicated by Southern blot analysis (Fig. 1). The S1a repeat contains one site each for *NdeI* (position 409–414) and *EcoRV* (position 466–471), and two sites for *KpnI* (positions 42–47 and 151–156). The S1b repeat contains one site for *NdeI* (position 248–253) and one for *EcoRV* (position 304–309).

A simplified version of the average genomic sequence is the “genomic consensus sequence” in which only the major variable positions are indicated using the IUBMB single letter code. We define a variable position of an average genomic sequence as “major” or “minor” depending on whether the percentage of the minor base reaches or not the 10% threshold in all sequence electropherograms (see Section 2.5). This conversion is necessary for submission of these sequences to GenBank and represents a simplified description of the structural variability of the S1a or S1b repeat in a genome. To

illustrate this variability, a 150-bp portion of the genomic consensus sequences of the S1a and S1b repeats from the eight population samples of *R. dalmatina* is compared with the corresponding part of the MCS of each repeat in Fig. 3A and B, respectively (GenBank accession nos. AJ543383–AJ543390 and AJ543391–AJ543398). The S1a and S1b genomic consensus sequences differ among the various samples only for the presence or absence of some major variable positions. In the eight population samples the major variable positions are 16–28 in the S1a repeat and 9–17 in the S1b repeat. Note that each major variable position always contains the same main and minor base in all population samples where it is present.

The overall structural variability of a repeat in the genome is defined by the “average genomic sequence” indicating the percentages of the main and minor base(s) present in each variable position. A short segment of the S1a genomic consensus sequences from the eight population samples tested is shown Fig. 4 to illustrate the general characteristics of these sequences. In about 80% of the sequence positions, the S1a and S1b average genomic sequences contain a unique and identical base in all samples. Most of the other positions contain two (rarely three)

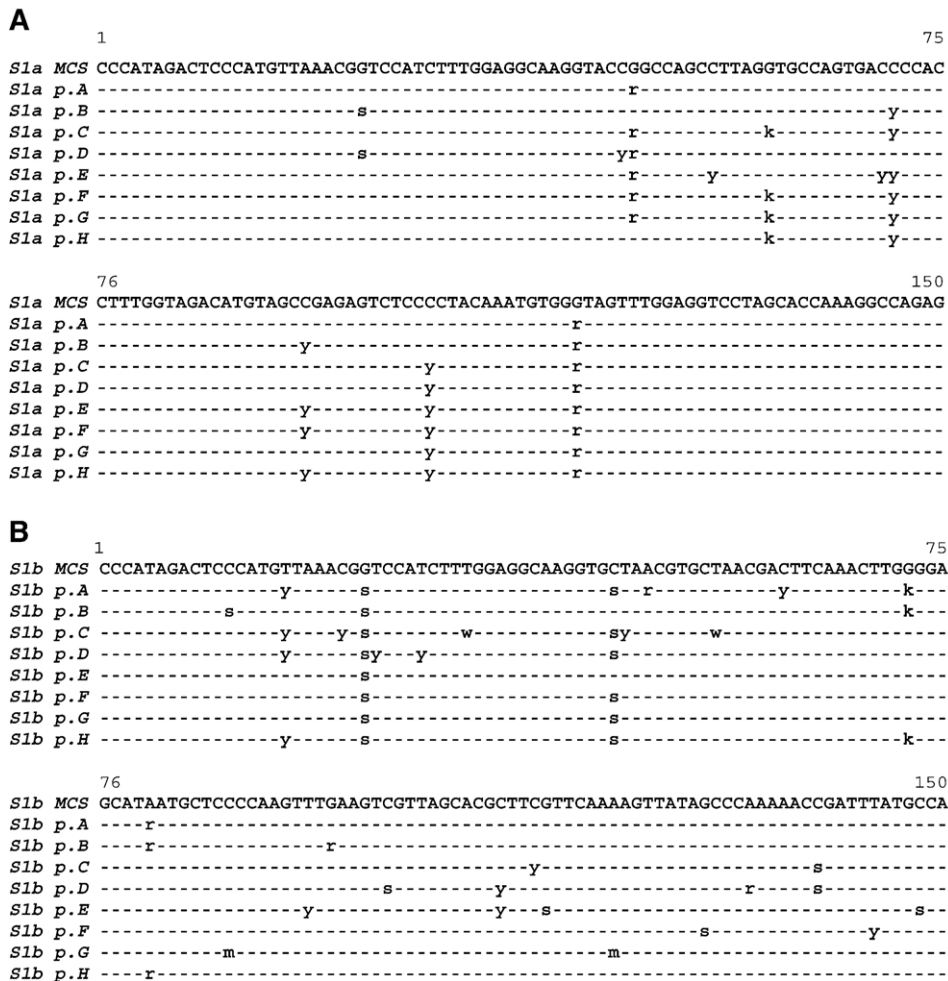


Fig. 3. Genomic consensus sequences of the S1a and S1b repetitive unit from *R. dalmatina*. A 150-bp portion (bp 1–150) of the genomic consensus sequences of the S1a repeat (A) or the S1b repeat (B) from population samples A to H of *R. dalmatina* (lines 2–9) is compared to the most common sequence of each repeat (line 1). Lower case letters indicate the major variable positions present in each population sample.

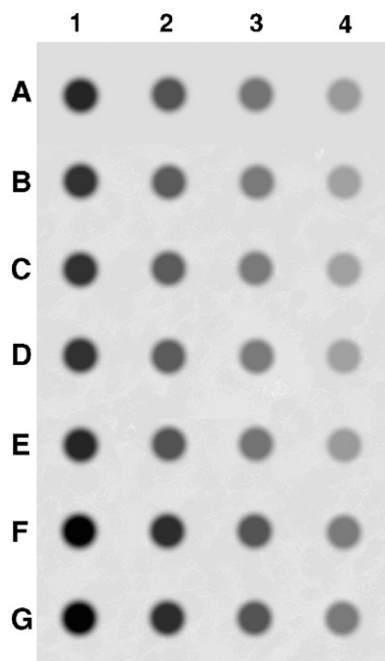


Fig. 6. Quantitative dot blot hybridization. Rows A–E: DNA from populations H, G, F, C and A of *R. dalmatina*, respectively; rows F and G, DNA from a population of *R. italica*, used as a reference. The amount of genomic DNA in the dots of columns 1–4 was 100, 50, 25, and 12.5 ng, respectively. Hybridization was carried out with a biotin-labelled probe of the S1a repeat from *R. dalmatina*.

number of S1a and S1b repeat variants whose variability is essentially restricted to the same base changes in specific positions of the same MCS.

3.3. Characterization of cloned repetitive units

We have cloned S1a and S1b repetitive units from genomic DNA of an Italian specimen. A 150 bp-long part of the sequence of five clones of S1a and S1b repeat is reported in Fig. 5A and B, respectively (for the complete sequences, see GenBank, accession nos. AJ504463–AJ504467 and AJ504468–AJ504472, respectively). As reference, the corresponding part of the MCS of each repeat is reported in the top line. We have chosen a portion of the sequence common to both repeats to show again their structural differences.

As expected, each repeat contains a different combination of the main to minor base substitutions from the MCS in the major (lower case letters) and minor variable positions (upper case letters) of the sequence. These base substitutions account for about 80% of all mutations from the MCS, but represent only 5% of all possible base substitutions. Eight of the ten clones examined also contain changes from the base(s) expected to be present in invariant or variable positions of the sequence (underlined upper case letters). The average frequency of these mutations is almost two orders of magnitude lower than that of the main to minor base substitution. This explains why only the latter base changes are detectable in the sequence electropherograms.

The structure of composite S1a–S1b repetitive units (Fig. 1) has been directly verified by cloning 0.70–0.85-kb fragments obtained from DNA digested with *Asp718I* (a *KpnI* isoschizomer). One 826-bp clone and three 717-bp clones were obtained (GenBank accession nos. AJ292000–AJ292003). These clones had exactly the size and sequence of a composite repeat formed by one S1a and one S1b complete repeat arranged in tandem, either intact (826-bp clone) or lacking the 109-bp *KpnI*–*KpnI* fragment of S1a (717-bp clones). In all clones, the S1a–S1b repeat junction occurs at the level of their homologous parts without solution of the sequence continuity.

3.4. Amount and chromosomal location of S1 satellite DNA

Quantitative dot blot analysis with probes of the S1a repeat from *R. dalmatina* (Fig. 6) or *R. italica* (not shown), shows that the molar content of the S1 satellite in DNA samples from various populations of *R. dalmatina* is $55 \pm 12\%$ of that present in the genomic DNA of *R. italica*, and corresponds to about 44 ± 10 fmol per microgram of DNA. This indicates that the S1 satellite DNA accounts for about 1% of the frog genome and that *R. dalmatina* has about $1.3 \pm 0.3 \times 10^5$ copies of S1a and S1b repeats per haploid genome assuming a genome size of 9.7 pg of DNA per nucleus (Odierna, 1989).

We also tested the chromosomal location of S1 satellite in *R. dalmatina* by FISH, using a biotinylated probe of the S1a repeat that hybridizes with both the S1a and S1b repeats. As illustrated in Fig. 7, the S1 satellite DNA is located in a pericentromeric position on one arm (chromosome pairs 3–5 and 9–13) or on both arms (chromosome pairs 1 and 2). This distribution is



Fig. 7. Chromosomal localization of the S1 satellite DNA of *R. dalmatina* by FISH. White bar represents 10 μ m.

completely different from that previously observed for the same satellite DNA in *R. italica* (Cardone et al., 1997) and *R. graeca* (Picariello et al., 2002).

4. Discussion

The present report provides one of the most detailed characterizations of a satellite DNA in a species presently available. In addition to the usual characteristics, we have determined two fundamental features not analyzed in studies on other satellite DNAs: the internal organization and the overall variability of repetitive units in the species.

The S1 satellite DNA from *R. dalmatina* is composed of two basic repeats S1a and S1b, like that from all other European brown frog species except *R. temporaria* that lacks the S1b repeat. The S1a repeats of *R. dalmatina*, *R. italica*, *R. graeca* and *R. temporaria* have the same size (494 bp), but different sequence (Picariello et al., 2002; Feliciello et al., 2005). By contrast, the S1b repeat is species-specific both in size and sequence being originated from the S1a repeat by a different deletion in each species. The S1b repeat from *R. dalmatina* has an additional peculiar feature: the initial part of the conserved S1a segment is duplicated in inverted orientation generating a 30 bp-long palindrome.

The chromosomal location and copy number per haploid genome of the S1 satellite DNA from *R. dalmatina* are also characteristic of the species. As discussed elsewhere, the species-specificity of all main features of S1 satellite DNA support the assumption that this repetitive DNA is an absolute molecular marker of species (Picariello et al., 2002).

The relevance of the present study is not due to the definition of these conventional features, but to the characterization of the organization and the overall sequence variability of the S1a and S1b repeats in the species. Indeed, the lack of meaningful population data is the main reason why the evolution and function of this repetitive DNA has been poorly understood up to now. This limitation is overcome in the S1 satellite DNA. Its high sequence and size homogeneity allowed us to sequence the mixture of repeats amplified from genomic DNA and to obtain average genomic sequences that represent the overall variability of the repetitive units present in a genome. The analysis of these sequences in specimens from different populations allows defining the overall variability of a repetitive unit in the species. Moreover, being the S1 satellite DNA present in various species, we could also compare its overall structural variability among these species.

We have recently carried out the same type of analysis on the S1a repetitive unit from *R. temporaria*, the European brown

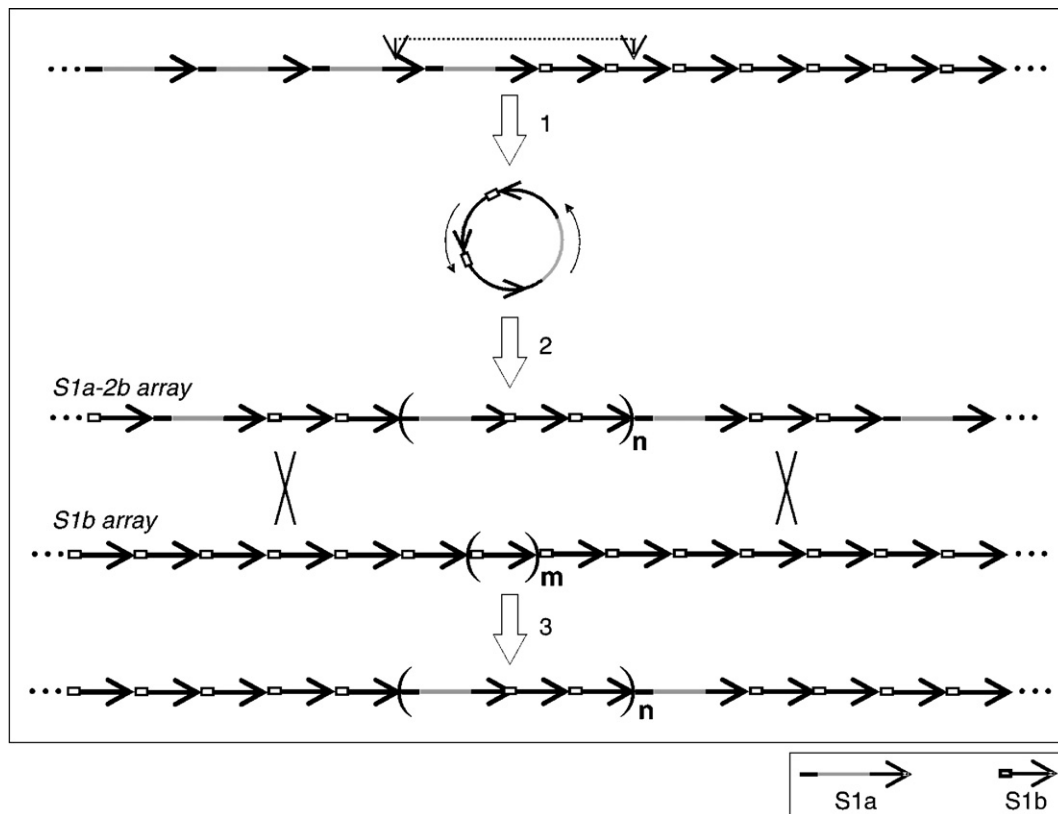


Fig. 8. A possible model of the origin of the composite S1a–S1b arrays in brown frogs. Step 1: intra-molecular homologous recombination at the junction of S1a and S1b arrays produces a circular DNA template containing complete copies of S1a and S1b repeats (e.g. one S1a and two S1b repeats). Step 2: rolling circle amplification generates a composite S1a–S1b array (e.g. S1a–2b array). Step 3: the composite S1a–S1b replaces part of an S1b (or S1a) array by homologous recombination. Because of the homology of S1a and S1b repeats, the steps 1, 2 and 3 can take place in S1a, S1b and composite S1a–S1b arrays as well, determining expansions and contractions of each type of array in the genome.

frog species with the largest geographic distribution (Feliciello et al., 2005). This study has shown that the S1a repeats display a unique, species-specific MCS in the whole species, and that repeat variability is essentially restricted to specific base changes in few positions of the MCS. Analysis of the average genomic sequences indicated the presence in each genome of thousands of S1a repeat variants with a different distribution of main and minor bases in these variable positions. Moreover, most of the S1a variants present in the genome of specimens from various European populations are different, showing the presence of an enormous number of repeat variants in the species. The data presented here confirm that the S1a and S1b repeat of *R. dalmatina* display exactly the same features. As discussed in detail in the previous report, crossing-over and gene conversion cannot account for this type of variability as postulated by current models of satellite DNA evolution.

Gene conversion is due to the repair of miss-matches in hetero-duplexes generated by migration of the recombination fork during chromosome recombination. According to current models, this is the main process responsible for concerted evolution of satellite DNA as a result extensive miss-match repair is expected to occur as a result of the sequence variability of repetitive units. In *R. dalmatina* most S1a and S1b repeats are joined in composite repetitive units without solution of the sequence continuity. This provides a unique opportunity to verify the validity of current models. Crossing-over in this satellite DNA should produce S1a–S1b hetero-duplexes frequently, and therefore gene conversion should eliminate sequence difference between the homologous parts of the two repeats. The fact that the sequence of S1b repeat is completely different in several positions from that of the S1a homologous counterpart demonstrates that no gene conversion and consequently no crossing-over occurs in this satellite DNA. This conclusion is consistent with available data indicating that crossing-over takes place in euchromatic regions of chromosomes in all higher organisms (for reviews see Brown, 1966; Resnick, 1987). Evidence that crossing-over occurs very rarely if at all in satellite DNA has been provided by Wevrick and Willard (1989) who could not detect any meiotic or mitotic recombination event in human α -satellite DNA arrays from six autosomal loci in three-generation families.

A high pressure of mutation acting on satellite DNA is indicated by the sequence variability of satellite DNA repetitive units and data showing that DNA repair takes place much less efficiently in heterochromatin than in euchromatin (Slijepcevic and Natarajan, 1994; Surrales et al., 1997). Our studies on the S1 satellite DNA of European brown frogs indicate the presence of a very large number of S1a and S1b repeat variants showing variability essentially restricted to specific base changes in a limited number of positions. Analysis of individual S1a and S1b repeats cloned from genomic DNA of *R. dalmatina* and other frog species confirms that most variability is due to the main to minor base changes in the variable positions, but also shows the presence in many repeats of some random mutations in the invariant positions. All these features are perfectly consistent with a high pressure of mutation acting on satellite DNA if a repair mechanism were present that periodically removes all

mutations except the main to minor base changes that are tolerated and allowed to accumulate in satellite DNA.

In some brown frog species (*R. italica*, *R. dalmatina*, *Rana pyrenaica* and *Rana arvalis*) the S1b repeat has no sites for one or more restriction enzymes present in the homologous S1a repeat. Digestion with an S1a-specific enzyme produces different satellite DNA fragments depending on whether the two repeats are organized as separate arrays or as composite S1a–S1b arrays. This is a unique feature, not found in any other known satellite DNA, which allows defining the overall organization of the two repeats in the genome of these species as well as eventual changes of this organization simply by Southern blot analysis.

Southern blots produced very surprising results in *R. italica*, the first species we have investigated. We found that the fraction of total repeats organized as composite S1a–S1b repetitive units varied from about 1% to 23% in six specimens of *R. italica* from a small stream in South Italy (Cardone et al., 1997). In this species the S1a and S1b repeats are mostly organized as distinct, adjacent arrays located on both arms of all chromosomes close to centromere. This means that two individuals from the same community present a different organization of 1.3×10^5 copies of S1a and S1b repeats distributed among many chromosomes. At that time we were unable to explain this astonishing observation because the processes that usually determine genetic variability in a community (crossing-over, independent chromosome assortment and sexual reproduction) not only cannot generate the composite S1a–S1b repeats, but also should rapidly distribute these repeats between homologous chromosomes and among individuals from the same community. After this report, we have repeatedly observed large variations of the amount of the composite repeat in individuals from the same community in other populations of *R. italica* and also of *R. arvalis* (unpublished results). All these observations indicate the presence of an unknown process that produces very rapid and extensive expansions and contractions of specific arrays in the genome.

In *R. dalmatina* we have found a predominant organization of the S1a and S1b repeat opposite to that present in *R. italica*, and defined the structure of composite S1a–S1b repetitive unit. Taken together, the characteristics the S1 satellite in the two species clearly point to a mechanism that uses rolling circle amplification and substitutions by homologous recombination to produce very rapid and extensive changes of satellite DNA arrays. As outlined in Fig. 8, this mechanism provides a simple explanation of the origin and structure of the composite S1a–S1b repeats as well as the different organization and behavior of the S1a and S1b repeats in diverse species. Assuming that the S1a and S1b arrays were originally contiguous in all brown frog species – as occurs in *R. italica* – the intra-molecular homologous recombination can generate circular templates either within each S1a or S1b array, or between homologous S1a and S1b repeats at the junction of the S1a and S1b arrays. In the latter case, rolling circle amplification generates composite S1a–S1b arrays that can replace S1a or S1b arrays by homologous recombination. Thus, the different organization of the S1a and S1b repeats in various frog species may simply

be the result of an original difference in the frequency of S1a–S1b array junctions, which is higher the shorter the length of the S1a and S1b arrays. In *R. dalmatina* (and *R. pyrenaica*) the presence of short S1a and S1b arrays could have initially favored a large accumulation of composite repeats, which then expanded to become the predominant type of organization. In *R. italica* and *R. arvalis*, the presence of long S1a and S1b arrays allows only a minor fraction of S1a and S1b units to be organized as composite repeats. In most *R. italica* specimens, this fraction corresponds to $5 \pm 2\%$ of the total S1a and S1b repeats, but values lower than 2% or higher than 10% are quite frequent. This indicates that fluctuations of the amplification–replacement events may occasionally determine the large reduction or increase of the composite repeats observed in *R. italica* and *R. arvalis*.

It is important to underline that the proposed mechanism could use either an extra-chromosomal or intra-chromosomal rolling circle amplification (Hourcade et al., 1973; Young and Cullum, 1987; Petit et al., 1992), and that Fig. 7 illustrates only one of the two modalities. In either case, the mechanism produces circular satellite DNA molecules, which indeed have been found in various organisms (Pont et al., 1987; Cohen et al., 1999). Walsh (1987) postulated that random and occasional events of rolling circle amplification are required to maintain long satellite DNA arrays in the genome. The amplifications occurring in the S1 satellite DNA do not match the Walsh's model at all, because these events are both frequent and highly directional. In fact, random amplifications would cause a rapid expansion in the genome of the base changes present in invariant positions of many repeats (Fig. 5), and cause a rapid loss of the satellite specific structure. By contrast, events that selectively amplify repetitive units carrying little or no mutations in the invariant positions and use the newly amplified arrays to replace satellite DNA segments rich in these mutations generate an overall structure of satellite DNA matching exactly that found in S1 satellite DNA. We conclude that these directional amplification–substitution events represent the action of a new DNA repair mechanism specific of satellite DNA.

Because of its tandem organization and presence on different chromosomes, satellite DNA is the genome component having the highest potential ability to cause recombination both between homologous and non-homologous chromosomes. This ability is blocked by the assembly in the constitutive heterochromatin, which is mediated by the binding of specific histone variants and non-histonic proteins (sat-proteins) (Harada et al., 1988; Fischer et al., 1994; Podgornaya et al., 2000; Malik and Henikoff, 2001; Talbert et al., 2002). All features of the substitution–repair mechanism point to its specific role in maintaining the ability of satellite DNA to bind specific sat-proteins and to assemble in heterochromatin. This is essential for karyotype stability as indicated by pathological conditions, like the ICF syndrome and many cancers, in which an altered heterochromatin condensation due to defective methylation of some satellite DNAs causes anomalous recombination events producing chromosome aberrations, (Tuck-Muller et al., 2000; Luciani et al., 2005; Wong et al., 2001; Ehrlich et al., 2003). The

model also implies a species-specific co-evolution of sat-proteins and their satellite binding sites as confirmed by recent studies on centromeric histone 3 variants in the *Drosophila* and *Arabidopsis* species-group (Malik and Henikoff, 2001; Talbert et al., 2002; Nagaki et al., 2003).

The repair mechanism determines the homogeneous structure of the repetitive units within the species, and the ability of satellite DNA to be constantly assembled in the constitutive heterochromatin, thereby blocking its great potential capability to participate to various recombination processes. Normally, these processes should occur rarely and have limited effects, like e.g. unequal crossing-over events between homologous chromosomes that may occasionally determine the array size variations observed in many satellite DNAs. However, when a mutation inactivates a sat-protein or modifies its DNA-binding specificity, the repair mechanism can rapidly replace the repetitive unit of the corresponding satellite DNA with new variants and sometime new satellite DNAs. During this transition, an altered condensation of heterochromatin at specific loci is expected to occur, and this should allow unusual recombination events that may produce relevant changes of the karyotype. These changes may include e.g. translocations due to recombination of satellite DNAs located on non-homologous chromosomes and changes of satellite DNA arrays produced by unequal crossing-over, insertions of circular satellite DNA, or deletions by intra-chromosomal recombination. The resulting karyotype changes should cause a reduction of fitness, and thus be readily eliminated from the species under conditions of stabilizing selection. However, under a strong selective pressure, mutations of sat-proteins in individuals carrying favorable characters could promote their reproductive isolation and the selection both of the advantageous characters and the modified karyotype, i.e. the formation of a new species.

Both effects are clearly evident in the S1 satellite DNA of brown frogs. The presence of a unique MCS indicates that the specific structure of each repeat was present and likely formed at the origin of the species, as also suggested by the fact that different deletion events originated the S1b repeat from the S1a repeat in each species. The presence of the same restricted variability of S1a and S1b repeats in all populations, even when separated by long distances and geographic barriers that prevent mutual genetic flow, indicates that the specific repeat structure has been maintained thereafter in each species by a non-stochastic mechanism. In high eukaryotes, each species usually presents a characteristic karyotype and species-specific satellite DNAs. Moreover, satellite DNAs of the same family are frequently found in related species, even in “living fossils” like sturgeons and cycads (de la Herran et al., 2001; Cafasso et al., 2003). All these features are consistent with the action of a mechanism that can determine relevant changes of the karyotype and satellite DNA structure during speciation, but also maintain the structure of satellite DNA as long as the species exists.

Preliminary data on the S1 satellite from *R. pyrenaica* indicate that the satellite DNA repair mechanism uses intrachromosomal rolling circle amplification. This amplification modality could explain the hierarchical organization and

the variations of the repeat periodicity and sequence in arrays from different chromosomes observed in some satellite DNAs. A recent study on alpha-satellite DNA (Schindelhauer and Schwarz, 2002) indicates that these features are determined by a “fast intrachromosomal conversion mechanism” that, in our opinion, corresponds to the repair mechanism described here.

In conclusion, the S1 satellite DNA of brown frogs presents an unusual combination of favorable features that has allowed the first characterization of the structural variability and internal dynamics of the repetitive units in a satellite DNA family. The definition of these two key properties provides direct and indirect experimental evidences that the structure and evolution of this satellite DNA family is mainly determined by a repair mechanism specific of this repetitive DNA. This mechanism not only provides a completely new model of satellite DNA evolution consistent with all available experimental data, but also indicates that this repetitive DNA plays a key role in the evolution of the species. For the first time, a precise molecular mechanism supporting the PE model of evolution proposed by Gould and Eldredge (1977) is available.

Acknowledgements

We thank Drs. Vincenzo Ferri (Milan, Italy), Henrik Bringsøe (Copenhagen, Denmark), Kurt Jørgensen, Wolfgang Böhme, and Alexander Kupfer (Zoologisches Forschungsinstitut und Museum Alexander König, Bonn, Germany), Martin Sandera (Znojmo, Czech Republic), Lennart Cederholm (Museum of Zoology, Lund University, Sweden), Sandro Tripepi (Università della Calabria, Cosenza, Italy), and Gaetano Odierna (Università Federico II, Naples, Italy) for providing *R. dalmatina* specimens and slides for FISH. We are grateful to Salvatore Cozzolino for automated sequencing, to Giuseppe Falcone (✉) for graphic assistance, and to Jean Ann Gilder for editing the manuscript.

References

- Bostock, C., 1980. A function for satellite DNA? Trends Biochem. Sci. 5, 117–119.
- Brown, S.W., 1966. Heterochromatin. Science 151, 417–425.
- Brutlag, D.L., 1980. Molecular arrangement and evolution of heterochromatic DNA. Annu. Rev. Genet. 14, 121–144.
- Cafasso, D., Cozzolino, S., De Luca, P., Chinali, G., 2003. An unusual satellite DNA from *Zamia paucijuga* (Cycadales) characterised by two different organisations of the repetitive unit in the plant genome. Gene 311, 71–79.
- Cardone, D.E., Feliciello, I., Marotta, M., Rosati, C., Chinali, G., 1997. A family of centromeric satellite DNA from the European brown frog *Rana graeca italica*. Genome 40, 774–781.
- Charlesworth, B., Sniegowski, P., Stephan, W., 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371, 215–220.
- Cohen, S., Menut, S., Méchali, M., 1999. Regulated formation of extrachromosomal circular DNA molecules during development of *Xenopus laevis*. Mol. Cell. Biol. 19, 6682–6689.
- de la Herran, R., et al., 2001. Slow rates of evolution and sequence homogenization in an ancient satellite DNA family of sturgeons. Mol. Biol. Evol. 18, 432–436.
- Dover, G., 1982. Molecular drive: a cohesive model of species evolution. Nature 299, 111–117.
- Ehrlich, M., et al., 2003. Satellite DNA hypomethylation in karyotyped Wilms tumors. Cancer Genet. Cytogenet. 141, 97–105.
- Feliciello, I., Chinali, G., 1993. A modified alkaline lysis method for the preparation of highly purified plasmid DNA from *Escherichia coli*. Anal. Biochem. 212, 394–401.
- Feliciello, I., Picariello, O., Chinali, G., 2005. The first characterisation of the overall variability of repetitive units in a species reveals unexpected features of satellite DNA. Gene 349, 153–164.
- Fischer, T.C., Groner, S., Zentgraf, U., Hemleben, V., 1994. Evidence for nuclear phasing and a novel protein specifically binding to cucumber satellite DNA. Z. Naturforsch. 49, 79–86.
- Gould, S.J., Eldredge, N., 1977. Punctuated equilibria: the tempo and mode of evolution reconsidered. Paleobiology 3, 115–151.
- Harada, M., Ouchi, K., Ohata, S., Kikuchi, A., Mizuno, S., 1988. Purification and characterization of W-protein. A DNA-binding protein showing high affinity for the W chromosome-specific repetitive DNA sequences of chicken. J. Biol. Chem. 263, 13952–13961.
- Henikoff, S., 2000. Heterochromatin function in complex genomes. Biochim. Biophys. Acta 1470, 1–8.
- Hourcade, D., Dressler, D., Wolfson, J., 1973. The amplification of ribosomal RNA genes involves a rolling circle intermediate. Proc. Natl. Acad. Sci. U. S. A. 70, 2926–2930.
- John, B., Miklos, G.L.G., 1979. Functional aspects of satellite DNA and heterochromatin. Int. Rev. Cytol. 58, 1–114.
- Luciani, J.J., et al., 2005. Subcellular distribution of HP1 proteins is altered in ICF syndrome. Eur. J. Hum. Genet. 13, 41–51.
- Malik, H.S., Henikoff, S., 2001. Adaptive evolution of Cid, a centromere-specific histone in *Drosophila*. Genetics 157, 1293–1298.
- Modi, W.S., 1993. Heterogeneity in the concerted evolution process of a tandem satellite array in meadow mice (*Microtus*). J. Mol. Evol. 37, 48–56.
- Nagaki, K., et al., 2003. Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres. Genetics 163, 1221–1225.
- Odierna, G., 1989. Localization of the nucleolar organizer and heterochromatin evolution in three European brown frogs: *Rana dalmatina*, *Rana graeca*, *Rana iberica*. Amphib.-Reptil. 10, 387–396.
- Petit, M.-A., Mesas, J.M., Noirot, P., Morel-Deville, F., Ehrlich, S.D., 1992. Induction of DNA amplification in the *Bacillus subtilis* chromosome. EMBO J. 11, 1317–1326.
- Picariello, O., Feliciello, I., Bellinello, R., Chinali, G., 2002. The S1 satellite DNA as a taxonomic marker in brown frogs: molecular evidence that *Rana graeca graeca* and *Rana graeca italica* are different species. Genome 45, 63–70.
- Podgornaya, O.I., Dey, R., Lobov, I., Enukashvily, N.I., 2000. Human satellite 3 (HS3) binding protein from the nuclear matrix: isolation and binding properties. Biochim. Biophys. Acta 1497, 204–214.
- Pont, G., Degroote, F., Picard, G., 1987. Some extrachromosomal circular DNAs from *Drosophila* embryos are homologous to tandemly repeated genes. J. Mol. Biol. 195, 447–451.
- Resnick, M.A., 1987. Investigating the genetic control of biochemical events in meiotic recombination. In: Moens, P.B. (Ed.), Meiosis. Academic Press, New York, pp. 157–210.
- Schindelhauer, D., Schwarz, T., 2002. Evidence for a fast, intrachromosomal conversion mechanism from mapping of nucleotide variants within a homogeneous alpha-satellite DNA array. Genome Res. 12, 1815–1826.
- Slijepcevic, P., Natarajan, A.T., 1994. Distribution of X-ray-induced G2 chromatid damage among Chinese hamster chromosomes: influence of chromatin conformation. Mutat. Res. 323, 113–119.
- Smith, G.P., 1976. Evolution of repeated sequences by unequal crossover. Science 191, 528–535.
- Stephan, W., 1989. Tandem-repetitive noncoding DNA: forms and forces. Mol. Biol. Evol. 6, 198–212.
- Surrales, J., Darroudi, F., Natarajan, A.T., 1997. Low level of DNA repair in human chromosome 1 heterochromatin. Genes, Chromosomes Cancer 20, 173–184.
- Talbert, P.B., Masuelli, R., Tyagi, A.P., Comai, L., Henikoff, S., 2002. Centromeric localization and adaptive evolution of an *Arabidopsis* histone H3 variant. Plant Cell 14, 1053–1066.

- Tuck-Muller, C.M., et al., 2000. DNA hypomethylation and unusual chromosome instability in cell lines from ICF syndrome patients. *Cytogenet. Cell Genet.* 89, 121–128.
- Walsh, J.B., 1987. Persistence of tandem arrays: implication for satellite and simple-sequence DNAs. *Genetics* 115, 553–567.
- Wevrick, R., Willard, H.F., 1989. Long-range organization of tandem arrays of α satellite DNA at the centromeres of human chromosomes: high-frequency array-length polymorphism and meiotic stability. *Proc. Natl. Acad. Sci. U. S. A.* 86, 9394–9398.
- Wong, N., et al., 2001. Hypomethylation of chromosome 1 heterochromatin DNA correlates with q-arm copy gain in human hepatocellular carcinoma. *Am. J. Pathol.* 159, 465–471.
- Young, M., Cullum, J., 1987. A plausible mechanism for a large-scale chromosomal DNA amplification in streptomycetes. *FEBS Lett.* 212, 10–14.