Available online at www.sciencedirect.com

SCIENCE DIRECT

Gene 349 (2005) 153–164

GENE
SECTION
EVOLUTIONARY GENOMICS

www.elsevier.com/locate/gene

The first characterisation of the overall variability of repetitive units in a species reveals unexpected features of satellite DNA

Isidoro Feliciello[a], Orfeo Picariello[b], Gianni Chinali[a,*]

[a]*Dipartimento di Medicina Clinica e Sperimentale, Facoltà di Medicina e Chirurgia, Università di Napoli Federico II, Via Pansini 5, I-80131 Napoli, Italy*
[b]*Dipartimento di Zoologia, Facoltà di Scienze MM. FF. NN., Università di Napoli Federico II, Via Mezzocannone 8, I-80134, Napoli, Italy*

Received 5 May 2004; received in revised form 21 October 2004; accepted 6 December 2004
Available online 14 March 2005
Received by M. Di Giulio

Abstract

We investigated the overall variability of the S1a satellite DNA repeats in ten European populations of *Rana temporaria* by a new procedure that determines the average sequence of the repeats in a genome. The average genomic sequences show that only 17% of the S1a repeat sequence (494 bp) is variable. The variable positions contain the same major and minor bases in all or many of the population samples tested, but the percentages of these bases can greatly vary among populations. This indicates the presence in the species of an enormous number of repeats having a different distribution of bases in these variable positions. Individual genomes contain thousands of repeat variants, but these mixtures have very similar characteristics in all populations because they present the same type of restricted and species-specific variability. Southern blots analyses and sequences of cloned S1a repeats fully support this conclusion. The S1 satellite DNA of other European brown frog species also presents properties indicating the same type of variability.

This first characterisation of the overall repeat variability of a satellite DNA in a species has revealed features that cannot be determined by gene conversion and crossing over. Our results suggest that a specific directional process based on rolling circle amplification should play a relevant role in the evolution of satellite DNA.

Keywords: Satellite DNA; Repeat variability; Concerted evolution; Rolling circle; *Rana*

1. Introduction

Highly repetitive DNA sequences organised in long direct tandem arrays are defined as satellite DNA and may constitute a relevant part of the genomic DNA of eukaryotic organisms. The large clusters of satellite DNA sequences are mostly located at centromere and telomeres of chromosomes, and represent a consistent part of the DNA present in the constitutive heterochromatin (Brutlag, 1980; Beridze, 1986).

So far, no defined function has been clearly demonstrated for this highly repetitive DNA, although it could play a role in the genomic structure and evolutionary processes (Bostock, 1980). A possible contribution of centromeric satellite DNA to the centromere function has been also suggested (Henikoff, 2000). Various models have been proposed for the origin and evolution of satellite DNA (e.g. Smith, 1976; Dover, 1982; Charlesworth et al., 1994), but unequivocal experimental evidence in favor of any of these models is lacking. This repetitive DNA shows very variable sequence and size of the repetitive unit among species (Brutlag, 1980), but similar satellite DNAs are frequently present in species of the same species-group or genus (Beridze, 1986).

A fundamental reason for our limited knowledge is that satellite DNA represents the portion of the eukaryotic genomes less accessible to structural analysis. Its organisa-

Abbreviations: bio-16-dUTP, biotin-16-2′-deoxy-uridine-5′-triphosphate; bp, base pair(s); kb, kilobase(s) or 1000 bp; Myr, million(s) of years; PCR, polymerase chain reaction; *R.*, *Rana*; sat-proteins, satellite DNA-binding proteins.

* Corresponding author. Tel.: +39 081 7463355; fax: +39 081 7463367.
E-mail address: chinali@unina.it (G. Chinali).

0378-1119/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2004.12.001

tion in long tandem arrays not only prevents the use of the clone overlapping and chromosome walking techniques needed to assemble long sequences, but also makes the analysis of recombination, amplification and substitution events in this DNA extremely difficult or impossible (Charlesworth et al., 1994).

The S1 satellite DNA of the European brown frogs was firstly characterised in *R. italica* (Cardone et al., 1997), and immediately presented features able to overcome one of the limitations usually met in the study of this DNA. The presence of two homologous repetitive units S1a (494 bp) and S1b (285 bp) in the same satellite DNA made possible to obtain evidence of specific processes that determine very large differences in the organisation of these repetitive units among frogs of the same population. In a more recent study (Picariello et al., 2002), we found that the mixture of the S1a or S1b repetitive units amplified by PCR from the genomic DNA of *R. graeca* could be directly sequenced to yield the average sequence of these two repeats in the genome.

The S1 satellite DNA from *R. temporaria* can be also investigated by this approach. Differently from *R. italica* and *R. graeca*, this brown frog has a very wide geographic distribution, being present in all of the Europe except for the southern parts of the Balkan Peninsula, Italy and Spain. The analysis of the S1 satellite DNA in various populations of *R. temporaria* has also been facilitated by the presence of only one type of repeat (S1a). The possibility to examine different populations living in distant regions separated by geographic barriers that limit or abolish a mutual genetic flow is fundamental to define the dynamics of a satellite DNA within a species.

In this report we have examined DNA samples from ten different European populations of *Rana temporaria* both by Southern blot analysis and by determining the average genomic sequence of the S1a satellite DNA repeat. Additional information has been obtained by the analysis of the sequence of the S1a repeat clones. Our report provides the first complete characterisation of the overall variability of a satellite DNA repeat within a species with a wide geographic distribution. This variability strongly suggests the presence of a specific directional process acting on this repetitive DNA, because it has features that cannot originate only from the action of stochastic processes as postulated by most current models of satellite DNA evolution.

2. Materials and methods

2.1. Animals

We analyzed specimens from ten populations of *R. temporaria* (A–J) of the following areas: Leon, Spain (A); Lourdes, France (B); Bristol, UK (C); Chamberry, France (D); Cuneo, Italy (E); Sondrio, Italy (F); Znojmo, Czech Republic (G); Copenhagen, Denmark (H); Lund, Sweden (I); Ekaterinenburg, Russia (J).

2.2. DNA isolation

DNA was always extracted from blood, skin or liver of single animals. Tissue homogenates were digested at 55 °C with proteinase K (0.2 mg/ml) in the presence of 0.5–1% sodium dodecyl sulphate before extraction with phenol, phenol-chloroform (1:1 v/v) and chloroform-isoamyl alcohol (25:1 v/v). After ethanol precipitation, samples were treated with RNase and proteinase K and processed as previously described (Cardone et al., 1997) to yield highly purified DNA. Results were independent of the tissue used for DNA preparation. DNA from clones of *R. temporaria* S1 satellite DNA in pTZ19R cloning 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 repeats were obtained by digestion of genomic DNA from *R. temporaria* with *Eco*RV. These fragments were cloned in the *Sma*I site of pTZ19R following a procedure previously described (Cardone et al., 1997).

2.4. PCR amplification

Because of the tandem organisation of satellite DNA, whole unit repeats of S1a from *R. temporaria* are amplified by PCR using primers with the same origin in the S1a sequence, but with opposite orientation. As previously reported, genomic DNA from single frogs was amplified with the two sets of primer pairs Rdr160–Rin161 and Rdr371–Rin372 and the resulting amplified S1a monomers were isolated by preparative gel electrophoresis (Picariello et al., 2002). The two sets of primers were chosen in two regions of the S1a sequence that are highly conserved in brown frogs, and have origins located at a distance of 210 bp in the sequence of the S1a repetitive unit.

2.5. DNA sequencing

After purification, the PCR-amplified S1a repeats were sequenced in both orientations by an automatic sequencer (Applied Biosystems, Foster City, Calif.) using the Big-Dye Terminators kit by the same manufacturer and the primers used for amplification. Each amplification product originates from a large number of similar S1a repetitive units and, as expected, electropherograms showed the presence of more than one base in several positions of the sequence. The direct and inverse sequence electropherograms of each amplified repeat were compared with Sequence Navigator software (Applied Biosystems): only bases present in both sequencing orientations were considered. The portion of sequence around each primer pair was determined by sequencing the amplification

product obtained with the other pair of primers. In each sequence electropherograms, the relative amount of each base in positions containing two or more bases was assumed to be proportional to the height of the corresponding signal. The percentage of a base in a given position of the sequence was calculated as the average of the percentages determined in the direct and inverse sequence electropherograms of both S1a amplified repeats, except for the region of each primer set in which only the sequence obtained with the other primer set could be used. The sequence resulting from the analysis of the two amplification products is defined "average genomic sequence". For submission to GenBank, the average genomic sequences were converted to a "genomic consensus sequence" using the IUBMB single letter code for multiple bases in the same position. A minor base was included in the genomic consensus sequence only if it gave a signal at least one-tenth that of the main base, both in the direct and inverse sequencing orientation.

The S1a monomers cloned in pTZ19R were sequenced using pUC18 direct and inverse primers. In the consensus sequence derived from pTZ19R clones, the IUBMB single letter code assignments were made when the same base change was present in at least two clones.

2.6. Biotin labeled probes

Biotinylated probes of S1a repeat were obtained by PCR using bio-16-dUTP (Sigma, St. Louis, U.S.A.) and purified S1a repeats amplified from genomic DNA of *R. temporaria* or *R. italica* as templates. The results obtained with the two types of probes were equivalent.

2.7. Southern blot and dot blot analyses

Southern blot and quantitative dot blot analyses were carried out as described (Picariello et al., 2002). Genomic DNA from a population of *R. italica* from Salerno province (Italy) containing 80 fmol of S1 satellite repeats per microgram was used as a reference.

3. Results

3.1. Southern blot analysis

The genome of European brown frogs contains the S1 satellite DNA family typical of this species-group. In *R. italica* and *R. graeca* the S1 satellite DNA presents two types of repetitive units, S1a and S1b, containing sites for the restriction enzymes *Kpn*I (*Asp*718I), *Eco*RV, *Nde*I, *Stu*I and *Nhe*I. The S1 satellite DNA of *R. temporaria* was digested by *Kpn*I, *Eco*RV, *Nde*I and, in part, by *Nhe*I, but not by *Stu*I (Fig. 1). In all cases, digestion produced fragments with a size of about 0.49 kb or multiples thereof, indicating that in *R. temporaria* the satellite is formed by

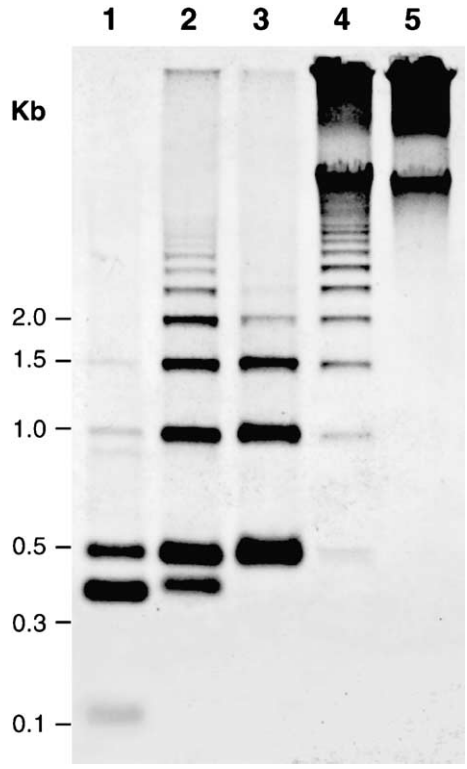

Fig. 1. Southern blot of the genomic DNA from *R. temporaria* digested with various restriction enzymes. Genomic DNA (3 μg) from a specimen of *R. temporaria* from population J (Russia) was digested with *Kpn*I (lane 1), *Eco*RV (lane 2), *Nde*I (lane 3), *Nhe*I (lane 4) and *Stu*I (lane 5). After electrophoresis on 1.5% agarose gel (1.5 h at 4 V/cm), samples were analyzed by Southern blot hybridisation with a biotin-labeled probe from the S1a satellite DNA of *R. italica*.

repeats of the S1a type. Like the corresponding repeats from other brown frogs, this repeat contains two *Kpn*I sites as indicated by the typical 0.38-kb and 0.11-kb fragments produced when the repeat is digested at both sites. The additional weak band of about 0.40 kb in the DNA digested with *Eco*RV likely suggests the presence of a second *Eco*RV site located about 90 bp from the first one in a minor fraction of the S1a repeats.

The same pattern of S1 satellite DNA bands illustrated in Fig. 1 for a specimen from Russia is observed when the DNA samples from all tested populations of *R. temporaria* are digested with *Kpn*I, *Nde*I or *Stu*I (data not shown). Digestion with *Eco*RV also produces the same pattern, including the presence of the 0.40-kb fragment in all ten populations (Fig. 2A). By contrast, the rare *Nhe*I restriction site is not found in every population tested (Fig. 2B) and is practically absent in specimens from the French side of Pyrenees and Spain.

Quantitative dot blot analysis shows that S1a satellite DNA content in the genome of specimens from various populations of *R. temporaria* is quite homogeneous and corresponds to 28±5 fmol per microgram of DNA (Fig. 3). This value indicates that the S1a satellite DNA accounts for 0.9% of the total genomic DNA of the frog and that about

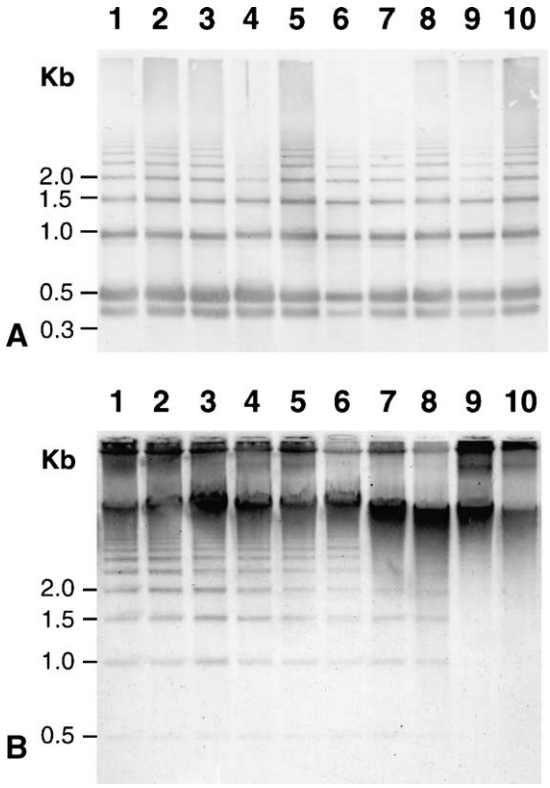

Fig. 2. Southern blots of genomic DNA from different populations of *R. temporaria*. Genomic DNA (3 μg) of specimens from populations A to J of *R. temporaria* was digested with *Eco*RV (A) or *Nhe*I (B) and analyzed by Southern blot hybridisation as described in Fig. 1.

0.9×10^5 copies of S1a repeat are present per haploid genome assuming a genome size of 10.2 pg of DNA/nucleus (Backmann and Nishioka, 1978).

3.2. Characterisation of the average genomic sequence of the S1a satellite DNA repetitive unit

Southern blots suggest a high homogeneity of the S1a repeats through the species. This homogeneity could be analysed experimentally. As in the case of *R. italica* and *R. graeca*, the mixture of S1a repeats amplified by PCR from genomic DNA of *R. temporaria* could be directly sequenced, although each amplification product contains thousands of different units simultaneously amplified by the PCR reaction. The resulting "average genomic sequence" is a representative sample of all the S1a repeats of the genome examined (Picariello et al., 2002).

We have determined the average genomic sequence of single frogs from ten different European populations. For submission to GenBank these sequences were converted to genomic consensus sequences in which only the major variable positions are indicated (see Section 2.5). A 125 bp portion of these sequences is shown in Fig. 4A to illustrate their features (GenBank accession numbers AJ515543–AJ515552). Variations among individuals from different European populations are limited to the presence or absence of minor bases in some positions of the consensus sequences. From these sequences, the most common sequence of the repeats present in the genome can be obtained by indicating the predominant base found in each variable position. A remarkable result is observed: all the ten population samples have the same most common sequence.

A general consensus sequence of the S1a repeat from *R. temporaria* can be obtained from these genomic consensus sequences (Fig. 4B). This sequence contains the restriction sites for *Kpn*I (positions 42–47 and 151–156), *Nde*I (positions 409–414) and *Eco*RV (positions 466–471), and twelve of the 13–21 major variable positions present in the genomic consensus sequences.

The species-specific common features of the S1a repeat are well described by these two general sequences. The characterisation of the repeat variability in the species requires a detailed analysis of the average genomic sequences from the ten populations tested. In Fig. 4C, a 25-bp segment of the 494-bp S1a repeat is chosen to illustrate the main features of these sequences. A comparison with the corresponding part of the genomic consensus sequences (Fig. 4A) makes evident that the "absence" of a base from a position of the consensus sequence usually corresponds to the presence of this base in the average genomic sequence in a percentage lower than the threshold limit. The analysis of these minor variable positions allows defining the positions of the less common *Eco*RV site and of the rare *Nhe*I site. The sequence electropherograms of the

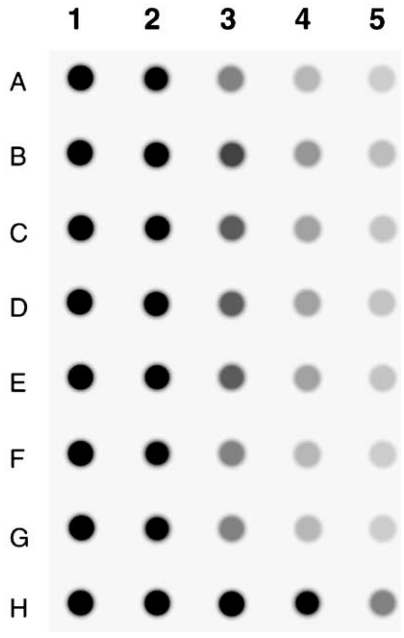

Fig. 3. Quantitative dot-blot hybridisation. Rows A to G, DNA from populations J, H, I, G, F, C and B of *R. temporaria*, respectively; row H, DNA from a population of *R. italica*. The amount of DNA in dots of columns 1–5 was 200, 100, 50, 25 and 12.5 ng, respectively. Hybridisation was carried out with a biotin-labeled probe of the S1a repeat from *R. temporaria*.

A

```
            370                                                                                                                         494
S1a M.C.S.  AACTGATAGCGACTTACTTTCTTCGCCACATTGTAGCCCCATATGTCCTGAACAGACCAAGCCAAGTTTGGGGTCCAAGGGTCCAATAGCCGTGGAGATATCGCGTCTCAAGCTAACCCTGTTTC
S1a pop.A   ---y-------------------y-----------------------------------------------------------------------------y-------------------w---
S1a pop.B   ------------------------------------------------------------------------------------------------------r-r--------------------
S1a pop.C   ------------------------------------------------------------------------------------------------------r-r--------------------
S1a pop.D   ---y-------------------yr----------------------------------------k------------------------------------r-r--------------------
S1a pop.E   ------------------------------------------------------------------------------------------y-----------r-r--------------------
S1a pop.F   ---y----k--------------yr----------------------------------------k-------------------------r----------r-r--------------------
S1a pop.G   --------------------------------------y---------------------------------------------------------------r-r--------------------
S1a pop.H   ---y---------------------------------------------------------------------------------------r----------r-r--------------------
S1a pop.I   --------k----------------------------------k---y----------w-------------------------------------------r-r--------------------
S1a pop.J   -----------------------------------------------------------------k------------------------------------r-d---m----------------
```

B

```
1                                             C                                                                              125
CCCATAGACTCCCATGTTAAACGGTCCATCTTTGGAGGCAAGGTACYGGCCAGCCTTAGGTGCCAGTGACCCCACCTTTGGTAGACATGTAGCCGAGAGTCTCCTCTACAAATGTGGGTAGTTTG

126                                                    T           A             A   AAAA T                                  250
GAGGTCCTAGCACCAAAGGCCAGAGGGTACCTAGTGTGACCGGTAAATCGGTAAAWTCCTGGTAAAARGTCCAAAAAACATHAAAWWWWTKATATGCTCCGGTTGCTATTCAGGAAGATGCTCCC

251      A                                                                                                                   375
CAAGTTTGARGGCTCTAGCACCTTCCGTTCAAAAGTTATAGCCCAAAAACCAATTTACACCAATTTCAGGCAGAAGTCCCAAAACACCAAATTTTGGATGGCCAGTTCTCCCGAACCATAACTGA

376                                                                                             G G                    494
TAGCGACTTACTTTCTTCGCCACATTGTAGCCCCATATGTCCTGAACAGACCAAGCCAAGTTTGGGGTCCAAGGGTCCAATAGCCGTGGAGATATCRCRTCTCAAGCTAACCCTGTTTC
```

C

	456																								480
M.C.S.	T	A	G	C	C	G	T	G	G	A	G	A	T	A	T	C	G	C	G	T	C	T	C	A	A
Pop.A	-	T2	A8	-	-	-	-	-	-	-	-	-	-	-	-	T13	A10	-	A9	-	-	-	-	-	-
Pop.B	-	-	A9	-	-	-	-	-	-	-	-	-	-	-	-	T5	A16	-	A15	-	-	-	-	-	-
Pop.C	-	T6	A6	-	-	-	-	-	-	-	-	-	-	-	-	T6	A30	-	A15	-	-	-	-	-	-
Pop.D	-	T5	A4	-	-	-	-	-	-	-	-	-	-	-	-	T8	A22	-	A14	-	-	-	-	-	-
Pop.E	-	T14	A6	-	-	-	-	-	-	-	-	-	-	-	-	T2	A15	-	A18	-	-	-	-	-	-
Pop.F	-	T6	A13	-	-	-	-	-	-	-	-	-	-	-	-	T4	A18	-	A25	-	-	-	-	-	-
Pop.G	-	T7	A2	-	-	-	-	-	-	-	-	-	-	-	-	T6	A16	-	A30T5	-	-	-	-	-	-
Pop.H	-	T6	A12	-	-	-	-	-	-	-	-	-	-	-	-	T8	A22	-	A33T4	-	-	-	A2	-	-
Pop.I	-	T6	A3	-	-	-	-	-	-	-	-	-	-	-	-	T7	A12	-	A15T4	-	-	-	A3	-	-
Pop.J	-	T5	A6	-	-	-	-	-	-	-	-	-	-	-	-	T7	A13	-	A16T14	-	-	-	A12	-	-

Fig. 4. Genomic sequences of the S1a repetitive unit from *R. temporaria*. All the S1a sequences are arbitrarily numbered starting from the origin of primer Rdr371. The main sites for the restriction enzymes *Kpn*I, *Nde*I and *Eco*RV are underlined; the rare sites for *Nhe*I and *Eco*RV are double underlined. (A) A 125 bp portion (bp 370–494) of the genomic consensus sequences of the S1a repeat from populations A to J of *R. temporaria* is reported. The most common sequence of the S1a repetitive unit (S1a M.C.S.) is indicated on top. The major variable positions of each genomic consensus sequence are indicated by one letter code lower case letters (e.g. *y=C+T*, *k=G+T*). (B) General consensus sequence of the S1a repeat of *R. temporaria* derived from the genomic consensus sequences of the ten population samples. The most common base present in each variable position is indicated on top. (C) A 25 bp portion (bp 456–480) of the average genomic sequences of the S1a repeat from populations A to J of *R. temporaria* is reported. The most common sequence (M.C.S.) present in all the populations is indicated on top. Where present, the minor bases and their percentages are indicated by the corresponding symbol and a number (e.g. A13=13% Adenine).

ten population samples show the presence of 8–13% of the minor base T in position 378 that determines the presence of this *Eco*RV in positions 374–379, at 92 bp from the main *Eco*RV site (see Fig. 4A). The rare *Nhe*I is located in positions 131–136 (see Fig. 4B) where the sequence CCTAGC contains traces of G (3% or less) in position 131 only in the sequence electropherograms of the samples showing this site. This location was confirmed by Southern blot analysis of S1a repeats amplified from specimens of Russia and Spain digested with *Nhe*I (data not shown).

In the ten population samples of *R. temporaria*, 75% of the positions of the average genomic sequences display only one specific base in all populations, 17% (83 positions) present the same major and minor base in 3–10 samples (e.g. positions 457, 458, 471 and 472 in Fig. 4C), while the residual positions show traces of minor bases only in one or two samples. In 65 of the 83 variable positions the same minor base is present in 6–10 population samples and in a few case a second minor base can also be present (e.g. position 474, Fig. 4C).

We have also examined the variability of the genomic sequence within a population. For this analysis we used six individuals of another population (Val di Susa, Turin province, Italy), because only 1–3 specimens were available for each of the ten populations previously examined. The average sequences of these individuals showed essentially the same variable positions, but the percentages of minor bases in some of these positions varied consistently thus determining similar, but not identical, genomic consensus sequences (GenBank accession numbers AJ830886–AJ830891).

The meaning of these characteristics becomes evident if one considers that the percent fraction of a base in a given position of the sequence corresponds to the fraction of total genomic repeats containing the base in that position. The average genomic sequence of each population sample examined indicates the presence of repeats containing an average of 5–6 minor bases distributed in about seventy variable positions. The varying percentages of minor bases in so many variable positions indicate that each genome contains thousands of repetitive units having a different distribution of the major and minor bases in the variable positions of the sequence. In the ten population samples, moreover, the percentage of the minor base varies in each position with little correlation even between bases located in adjacent positions and in many variable positions the minor base is not found in all populations. All these features indicate that an extremely large number, likely billions, of repeat variants are present in the species and that only a fraction of these variants is found in each population.

3.3. Characterisation of cloned repetitive units

In Fig. 5, the sequences of five repetitive units cloned from an Italian specimen of *R. temporaria* (GenBank accession numbers AJ305283–AJ305287) and the corresponding consensus sequence are compared with the most common sequence of the S1a repeat. The clones contain fifty-eight base substitutions, one insertion and one deletion with respect to the most common sequence. The major to minor base substitutions in the variable positions of the S1a sequence represent less than 6% of all the possible base substitutions, but account for 75% of all base changes. Moreover, all repeats show different combinations of these base substitutions, and the consensus sequence derived from cloned repeats has variable positions also present in the average genomic sequence. Thus, the features of cloned S1a repeats correspond exactly to those expected from the average genomic sequence.

3.4. Species-specific structure of the S1a satellite repeats

Southern blot analyses show that the S1b repeat has a different size in each European brown frog species, while the S1a repeat differs among species only for the presence or the absence of specific restriction sites (Chinali et al., 1999). This demonstrates that the S1b repeat is species-specific and suggests that the same could be true also for the S1a repeat.

We have previously characterised the genomic sequences of the S1a and S1b repetitive units in four population samples of *R. graeca* and that of the S1a repeat in three population samples of *R. italica* (Picariello et al., 2002). These genomic sequences clearly indicate that these repeats have a restricted variability and a species-specific most common sequence like the S1a repeat of *R. temporaria*. As illustrated in Fig. 6, these two features determine a complete differentiation of the S1a repetitive units of these three species. The most common sequence of the S1a repeat of *R. temporaria* presents 24 and 29 bases different from that of the corresponding repeat of *R. italica* and *R. graeca*, respectively. Analysis of the average genomic sequences indicates the complete absence of any common base in 8 and 10 of these positions, respectively. The most common sequences of *R. italica* and *R. graeca* differ in 26 positions, and have no common base in 8 of these positions (not shown). Thus, the three species have completely distinct populations of S1a repeats, because even the absence of a common base in a single position is sufficient to differentiate two populations of repeats. This is shown by base changes in positions 262, 264 and 375 (Fig. 6) that determine the exclusive presence of a *Stu*I site and the two *Nhe*I sites in the S1a repeat of *R. italica* and *R. graeca*, respectively (Picariello et al., 2002).

This species-specificity and restricted variability is also evident in the 19 clones of S1a and/or S1b repeat of *R. italica* and *R. graeca* previously characterised. The sequences of these cloned repeats are more than 96% identical to their respective most common sequence, and contain 114 base changes, 86% of which are major to minor base changes in the variable positions of the sequence.

We have also determined and deposited in GenBank the genomic sequences and the sequence of clones of the S1a and/or S1b repetitive unit of most other European brown

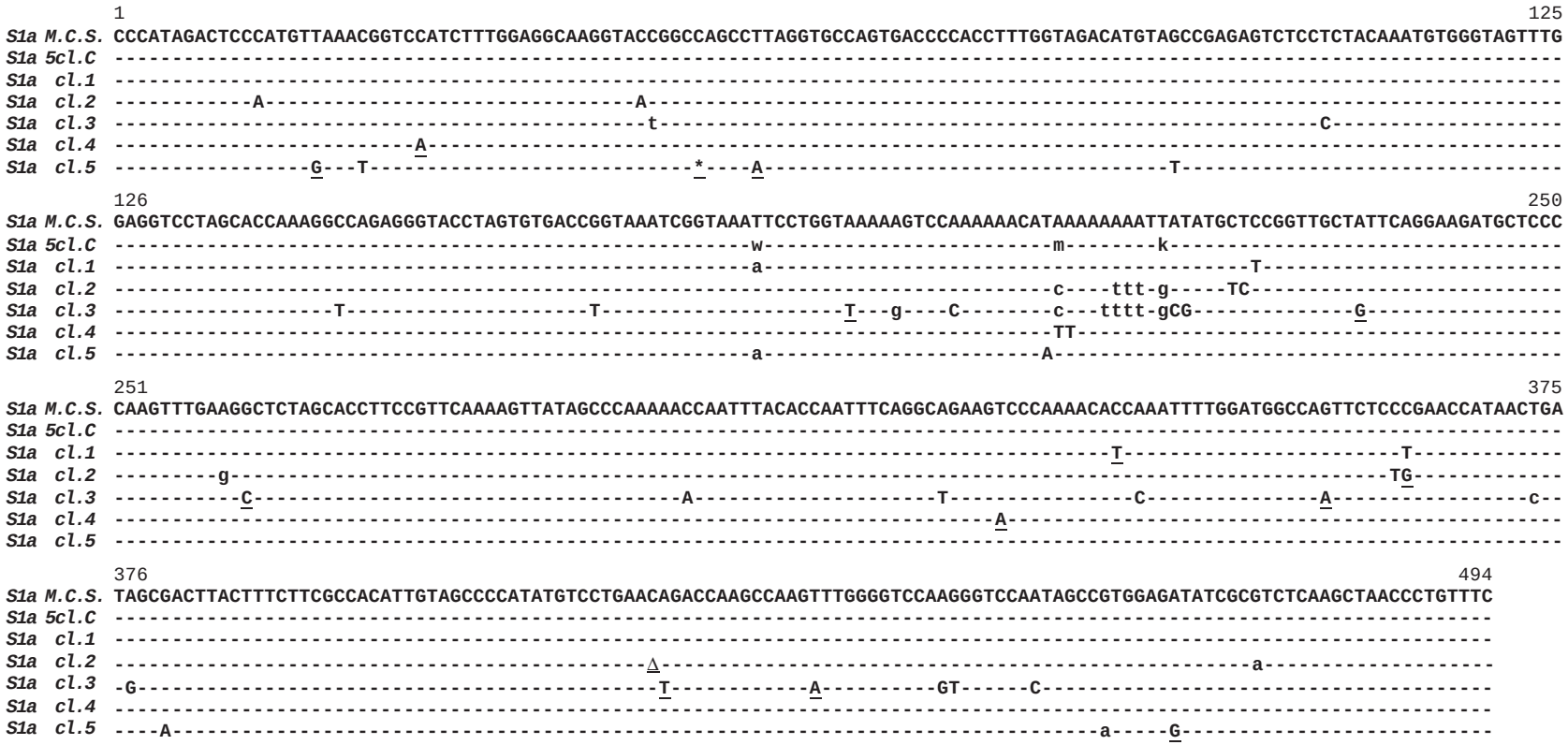

Fig. 5. Comparison of the sequences of cloned S1a repetitive units from *R. temporaria* with the most common sequence of the S1a repeat. The sequence of five clones of the S1a repeat (lines 3–7) from the DNA of a frog from population F (Lombardy, Italy) is compared with that of most common sequence of the S1a repeat (line 1). The consensus sequence derived from the five clones (S1a 5cl. C) is reported in line 2. Lower case and not underlined capital letters=minor base present in major and minor variable positions of the corresponding genomic consensus sequence, respectively; underlined capital letters=base substitution in an invariant position; Δ=one base deletion; *=one base insertion.

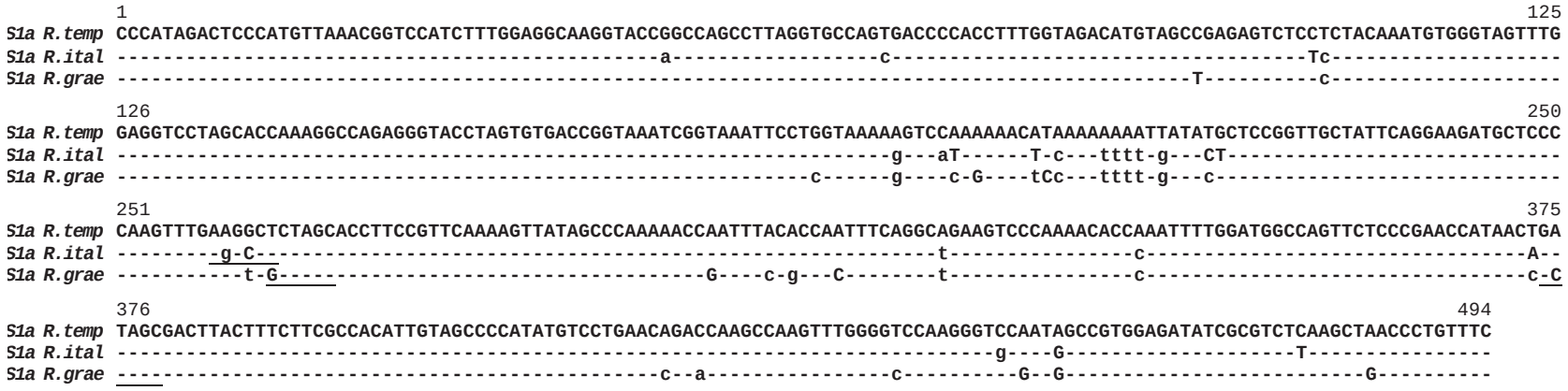

Fig. 6. Structural differences between the S1a repeat of *R. temporaria* and the S1a repeats of *R. italica* and *R. graeca*. The most common sequence of the S1a repeat of *R. temporaria* is compared to that of the same repeat of *R. italica* and *R. graeca*. In positions presenting a different base, an analysis of the average genomic sequences of the two compared species was carried out to verify the eventual presence in a species of a minor base corresponding to the major or minor base present in the same position in the other species. Lower case and capital letters indicate positions in which a common base is present or absent, respectively.

frog species (see e.g. *R. dalmatina*, GenBank accession numbers AJ504463–AJ504472 and AJ543383–AJ543398). As expected, all these sequences univocally indicate that these repeats have a restricted variability and a species-specific most common sequence like those of *R. temporaria*, *R. italica* and *R. graeca*.

4. Discussion

In this report we have investigated the overall structure of the repetitive unit of the S1 satellite DNA of *R. temporaria* in specimens from different European populations by Southern blot analysis and a new PCR-based method that determines the average sequence of the satellite DNA repeats in the genome.

The determination of the average genomic sequence of the S1a satellite repeat in many population samples of *R. temporaria* represents the most original and important contribution of this report. For the first time a precise characterisation of the overall intra-specific variability of a satellite DNA in a species with a wide geographic distribution becomes available. As described in a previous report (Picariello et al., 2002), the mixture of repeats amplified from genomic DNA by PCR using primers located in conserved parts of the repeat sequence is a highly reproducible and representative sample of the whole population of repeats present in the genome. This property has been further confirmed in this report by the ability of the average genomic sequences to determine the position and the relative frequency in the S1a repeat of all the restriction sites indicated by Southern blot analysis, including the rare additional *Eco*RV site and the very rare *Nhe*I site. Our results provide a detailed definition of the overall structure and variability of the S1a repetitive units in *R. temporaria*. The average genomic sequences indicate that the most common sequence of the S1a repeat is the same in all the population samples examined, and that the species contains a very large number of repeat variants that differ from this most common sequence mainly for the number and distribution of specific base changes in a limited number of positions of the S1a repeat. As expected from these features, each cloned repetitive unit corresponds to a different variant in which most or all base changes from the most common sequence are represented by major to minor base substitutions in these variable positions. This restricted, non-random variability is responsible for the homogeneity of the mixtures of repeat variants in individuals from different populations of the species in spite of the fact that in these individuals most repetitive units have a different sequence. The percentages of minor bases in the variable positions of the average genomic sequences show consistent variations even among individuals from the same population and, moreover, some variable positions are found only in some population samples. These features indicate that each genome contains only a fraction of the repeat variants present in a population and that each population contains only a fraction of the enormous number of repeat variants present in the species.

As described in Section 3.4, all the available sequence data, either published or deposited in GenBank, indicate that the S1a and S1b repetitive unit of the other European brown frog species have the same characteristics of the S1a repeat of *R. temporaria* (Cardone et al., 1997, Picariello et al., 2002). As illustrated in the example of Fig. 6, the restricted variability and the presence of a unique most common sequence of the S1a and S1b repeats in each species provide a precise molecular basis for the complete species-specificity of the S1 satellite DNA previously indicated by Southern blot analyses (Chinali et al., 1999). Differently from most other European species, *R. temporaria* has populations that – like the one from England – have been fully isolated from other European populations since the last glacial period. The presence of S1a repeats with the same characteristics in all populations of this species shows that the intra-specific homogeneity of the S1 satellite DNA is not due to genetic flow among populations.

Our data also provide indications about the origin and evolution of this satellite DNA. The fact that the most common sequence of S1a and S1b repetitive unit is the same in all populations of a species suggests that all repeats of a species have a common origin, and likely derive from the repeats present in the ancestors of the species.

Current models postulate that evolution of satellite DNA is mainly determined by gene conversion and crossing over, two processes associated with homologous recombination. However, a simple analysis makes evident that these processes can hardly account for the general features of the S1 satellite DNA of brown frogs.

These two processes should determine the fixation of different satellite DNA variants in related species. However, it is not easy to understand how processes based on homologous recombination could have fixed and could maintain two distinct homologous repetitive units in the same genome. Indeed, the sequence of the S1b repeat is practically identical to that of the corresponding part of the S1a repeat and, therefore, these two repeats should recombine freely.

The fact that each of the millions or billions of repeat variants found in a species has a different distribution of the same bases in the same variable positions is even more difficult to explain. According to current models, this variability should originate from different events of gene conversion and/or crossing over that occurred in each variant, but this would require an enormous frequency of recombination. By contrast, all experimental evidences indicate that recombination between homologous chromosomes occurs mainly in the euchromatin and is rare or undetectable in the constitutive heterochromatin where satellite DNA is located (for reviews see Brown, 1966; Resnick, 1987). The same appears to be true for the recombination of sister chromatids during mitosis (Bostock

and Christie, 1976). The meiotic and mitotic stability of satellite DNA has been directly confirmed by a study of Wevrick and Willard (1989) showing no recombination event in human α-satellite DNA arrays from six autosomal centromeres in three-generation families.

The remarkable variability of repetitive units indicates a strong tendency of satellite DNA to accumulate mutations, as also supported by experimental evidences and theoretical considerations. DNA mutations are corrected much less efficiently in heterochromatin than in euchromatin presumably because heterochromatic DNA is less available to the repair machinery of the cell (Slijepcevic and Natarajan, 1994; Surrales et al., 1997). Moreover, while mutations in functional or structural genes are frequently eliminated by natural selection, the same is not possible for a highly repetitive DNA. Satellite DNA is a highly repetitive DNA located in constitutive heterochromatin and, therefore, should represent the part of the genome that accumulates mutations at the highest rate. In our opinion, current models of evolution do not really explain how the rare events of gene conversion and unequal crossover occurring in satellite DNA could maintain this repetitive DNA homogeneous and, at the same time, allow a large accumulation of mutations.

Our data suggest a possible answer to this intriguing question. The presence of a very large number of variants with a restricted variability could indicate that accumulation of mutations occurs, but only at specific positions of the repetitive unit. This points to the action of a selective mechanism that eliminates most mutations from satellite DNA, except some defined base changes that are tolerated and allowed to accumulate in satellite DNA. This hypothetical selection mechanism is expected to maintain the ability of the repetitive DNA to assemble in the constitutive heterochromatin, as this is the only known feature common to all satellite DNAs. This feature is mediated by the binding of specific histonic and non-histonic proteins (sat-proteins) (Masumoto et al., 1989; Harata et al., 1988; Podgornaya et al., 2000; Malik and Henikoff, 2001; Talbert et al., 2002) and, therefore, this putative mechanism could have the function to maintain in the genome only the repetitive units that interact correctly with sat-proteins. In agreement with this prediction, recent reports have shown that the sequence of the satellite DNA binding site for some sat-proteins is rigidly maintained in the species. The human α-satellite DNA contains two types of repetitive unit, A and B, which bind the proteins pJ-α and CENP-B at the same site, respectively (Alexandrov et al., 2001). In each subfamily, the sequence of the binding site is highly conserved and specific for the corresponding protein suggesting the action of a selection mechanism that maintains the sequence and function of this binding site (Romanova et al., 1996). Recent studies on centromeric histone 3 variants in the *Drosophila* and *Arabidopsis* species-groups have also indicated a species-specific coevolution of these histone variants and the sequence of their satellite DNA binding site (Malik and Henikoff, 2001; Talbert et al., 2002; Nagaki et al., 2003).

It is evident that the selection mechanism suggested by these observations cannot be based on gene conversion and/or crossing over. Walsh (1987) postulated that the persistence of long satellite DNA arrays in the genome requires events of rolling circle amplification followed by the insertion of amplified repeats in the genome by homologous recombination. However, only indirect evidence of rolling circle amplification can be obtained in satellite DNA because the organisation of nearly identical repeats in long tandem arrays hinders the detection of any amplification–recombination event (Rossi et al., 1990). This is likely the reason why this process has been essentially ignored and considered to play a marginal role in satellite DNA evolution.

The S1 satellite DNA of some European brown frog species represents the only known exception to this rule, and has provided direct evidence that rolling circle replication occurs frequently in satellite DNA. Digestion of genomic DNA with a restriction enzyme specific of the S1a repeat produces different fragments when the S1a and S1b repeats are organised as distinct arrays (*R. italica* and *R. arvalis*) or as composite S1a–S1b repetitive units (*R. dalmatina*) (Chinali et al., 1999). We have previously reported that *R. italica* contains an extremely variable fraction of composite S1a–S1b repeats (Cardone et al., 1997). In six specimens of *R. italica* collected from a small river in South Italy, this fraction was found to vary from 1% to 23%, indicating that about 125,000 repeats distributed on most chromosomes have a completely different organisation in the genome of two frogs from the same community. This extraordinary variability cannot be produced by crossing over. Among known mechanisms, only rolling circle could determine rapid expansions or contractions of specific repeat subpopulations by producing amplified arrays that can replace segments of satellite DNA with a different organisation by homologous recombination. The very large variations observed also indicate that these events should occur rather frequently in the S1 satellite DNA.

Only repeats containing no or very few mutations in invariant positions of the sequence should be amplified. These mutations are present in many repeats and a random amplification would cause a large increase of the frequency of these mutations and determine the loss of the restricted variability of satellite DNA. The possibility that specific amplification–recombination events are part of a selection mechanism acting on satellite DNA is evident. Our hypothesis is that only repeats that can bind specific sat-proteins and assemble properly in heterochromatin are amplified, and that the resulting arrays are used to replace altered segments of satellite DNA. According to this model, only repetitive units containing mutations or combinations of mutations not affecting the binding of sat-proteins are maintained and eventually amplified in the genome. This should cause an accumulation of specific base changes only

at defined positions of the repetitive unit, thus determining the presence in the species of a very large number of repeat variants with a restricted variability.

This new model of satellite DNA evolution is also expected to determine a series of general features that are all confirmed by or consistent with the data available in the literature. The intra-specific variability of a satellite DNA should be mainly determined by the structural constraints imposed by sat-proteins, rather than by its ancient or recent origin. Moreover, this variability should not be distributed at random in the sequence of the repetitive units, even in the less homogeneous satellite DNAs. Indeed, the two most ancient satellite DNA families so far characterised in sturgeons (over 100 Myr old) and in the plant family of Zamiaceae (about 60 Myr old) present different intra-specific homogeneity (de la Herran et al., 2001; Cafasso et al., 2003), but are both much more homogeneous than the α-satellite DNA family of primates (about 30 Myr old) and less homogeneous than the S1 satellite DNA family of brown frogs (about 7 Myr old). Only very partial data are presently available on the intra-specific variability of most other satellite DNAs except for the human α-satellite DNA (for a review see Willard, 1998). A large number of repeats of this satellite DNA have been cloned, and their sequence was found to differ up to 35% within the same array and up to 50% between different arrays. A sequence analysis of 293 cloned monomers representative of most α-satellite DNA subfamilies originating from every human chromosome indicates that most monomers (80–97%) contain the same base in 75% of the repeat positions, while in 15% and 8% of the positions the variability is essentially limited to one and two specific base changes, respectively (Choo et al., 1991). The similarity of this type of variability with that present in the average genomic sequences of the S1 satellite repeats is evident in spite of the very low homogeneity of this satellite DNA. The frequent presence of the same base changes in two or more cloned repeats also suggests a non-random variability in other less characterised satellite DNAs.

The most relevant and original feature foreseen by the new model is the saltatory mode of evolution of satellite DNA. Mutations that change the binding specificity of a sat-protein are expected to determine a very rapid expansion of a new type of repetitive unit in the genome with concomitant extensive changes of the distribution and organisation of satellite DNA. Under conditions of stabilizing selection, mutations of sat-proteins should be rapidly eliminated from the species because individuals carrying these mutations should have a reduced fitness. However, under conditions of divergent or oriented selection, the same changes occurring in individuals carrying advantageous characters could favor their reproductive isolation and, thus, greatly accelerate the formation of a new species. Therefore, new satellite DNAs should mainly appear during formation of new species and, thereafter, they should maintain the same characteristics as long as the species exists. The presence of repetitive units with the same structure in all populations of the species and of different homologous satellite DNAs in species belonging to the same genus or species-group even in those of very ancient origin is fully consistent with this model (Garrido-Ramos et al., 1995; de la Herran et al., 2001; Cafasso et al., 2003).

The combined action of stochastic and directional processes that determine the fixation and maintenance of repetitive DNA variants in the species has been defined "molecular drive" by Dover (1982). Our studies on the S1 satellite DNA family of European brown frogs indicate that molecular drive causing the concerted evolution of satellite DNA should be mainly determined by a directional process based on rolling circle replication rather than by gene conversion and/or crossing over, as postulated by current models.

The structure, organisation and chromosomal distribution of satellite DNAs are some of the most relevant features that distinguish the genome of related species. A process that controls satellite DNA evolution should also play a specific and relevant role in the evolution of the species.

Acknowledgements

We thank Franco Andreone (Turin, Italy), Roger Avery (Bristol, UK), Henrik Bringsöe (Copenhagen, Denmark), Martin Sandera (Znojmo, Czech Republic), Trent Garner (Zürich, Switzerland), Vladimir Ischenko (Ekaterinenburg, Russia), Claude Miaud (Chamberry, France), Björn Lardner (Lund, Sweden), Gaetano Odierna (Naples, Italy) and Oscar Arribas (Barcelona, Spain) that kindly provided tissue specimens. We are grateful to Salvatore Cozzolino for automated sequencing and to Mr. Giuseppe Falcone for graphic assistance.

References

Alexandrov, I.A., Kazakov, A., Tumeneva, I., Shepelev, V., Yurov, Y., 2001. Alpha-satellite DNA of primates: old and new families. Chromosoma 110, 253–266.

Backmann, K., Nishioka, M., 1978. Genome size and nuclear size in Palaearctic frogs (*Rana*). Copeia 1978, 225–229.

Beridze, T., 1986. Satellite DNA. Springer Verlag, Berlin.

Bostock, C., 1980. A function for satellite DNA? Trends Biochem. Sci. 5, 117–119.

Bostock, C.J., Christie, S., 1976. Analysis of the frequency of sister chromatid exchange in different regions of chromosomes of the kangaroo rat (*Dipodomys ordii*). Chromosoma 56, 275–287.

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.

Chinali, G., Picariello, O., Feliciello, I., 1999. Species-specific structure and organization of the S1 satellite DNA family in the European brown frogs. In: Valakos, E., Lymberakis, P., Pafilis, P., Mylonas, M. (Eds.), Abstracts 10th Ordinary General Meeting of Societas Europaea Herpetologica. Natural History Museum of Crete, Irakleio, Greece, pp. 45–47.

Choo, K.H., Vissel, B., Nagy, A., Earle, E., Kalitsis, P., 1991. A survey of the genomic distribution of alpha satellite DNA on all the human chromosomes, and derivation of a new consensus sequence. Nucleic Acids Res. 19, 1179–1182.

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.

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.

Garrido-Ramos, M.A., Jamilena, M., Lozano, R., Ruiz-Rejón, C., Ruiz-Rejón, M., 1995. The *Eco*RI centromeric satellite DNA of the Sparidae family (Pisces, Perciformes) contains a sequence motive common to other vertebrate centromeric satellite DNAs. Cytogenet. Cell Genet. 71, 345–351.

Harata, 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.

Malik, H.S., Henikoff, S., 2001. Adaptive evolution of Cid, a centromere-specific histone in *Drosophila*. Genetics 157, 1293–1298.

Masumoto, H., Masukata, H., Muro, Y., Nozaki, N., Okazaki, T., 1989. A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. J. Cell Biol. 109, 1963–1973.

Nagaki, K., Talbert, P.B., Zhong, C.X., Dawe, R.K., Henikoff, S., Jiang, J., 2003. Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres. Genetics 163, 1221–1225.

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.

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.

Romanova, L.Y., et al., 1996. Evidence for selection in evolution of alpha satellite DNA: the central role of CENP-B/pJ alpha binding region. J. Mol. Biol. 261, 334–340.

Rossi, M.S., Reig, O.A., Zorzopulos, J., 1990. Evidence for rolling-circle replication in a major satellite DNA from the South American rodents of the genus *Ctenomys*. Mol. Biol. Evol. 7, 340–350.

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.

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., Comaiand, L., Henikoff, S., 2002. Centromeric localization and adaptive evolution of an *Arabidopsis* histone H3 variant. Plant Cell 14, 1053–1066.

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.

Willard, H.F., 1998. Centromeres: the missing link in the development of human artificial chromosomes. Curr. Opin. Genetics Dev. 8, 219–225.