Telomeric localization of the vertebrate-type hexamer repeat, $(TTAGGG)_n$, in the wedgeshell clam *Donax trunculus* and other marine invertebrate genomes*

Miroslav Plohl[a], Eva Prats[a], Andrés Martínez-Lage[b], Ana González-Tizón[b], Josefina Méndez[b] and Luis Cornudella[a]

[a] Department of Molecular and Cell Biology†, Institut de Biologia Molecular de Barcelona, CSIC, E-08034 Barcelona, Spain

[b] Department of Molecular and Cellular Biology, Universidade da Coruña, E-15071 A Coruña, Spain

Journal of Biological Chemistry, volume 277**,** issue 22, pages 19839–19846, 31 may 2002
Received january, 31, 2002; accepted march, 18, 2002; first published march, 20, 2002.

How to cite:

Telomeric localization of the vertebrate-type hexamer repeat, $(TTAGGG)_n$, in the wedgeshell clam *Donax trunculus* and other marine invertebrate genomes. Miroslav Plohl, Eva Prats, Andrés Martínez-Lage, Ana González-Tizón, Josefina Méndez and Luis Cornudella, *J Biol. Chem.*, 277(22), 2002, 19839–19846. DOI: 10.1074/jbc.M201032200

Abstract

The hexamer repeat sequence $(TTAGGG)_n$, found at the ends of all vertebrate chromosomes, was previously identified as the main building element of one member of a*Hin*dIII satellite DNA family characterized in the genome of the bivalve mollusc *Donax trunculus*. It was also found in 22 perfect tandem repeats in a cloned junction region juxtaposed to the proper satellite sequence, from which the DNA tract encompassing the clustered tandem copies was excised and subcloned. Here, the chromosomal distribution of $(TTAGGG)_n$ sequences in the *Donax* was studied by the sensitivity to *Bal*31 exonuclease digestion, fluorescence *in situ* hybridization (FISH) on metaphase chromosomes and rotating-field gel electrophoresis. To verify the occurrence of the hexamer repeat in the genomes of taxonomically related molluscs and other marine invertebrates, genomic DNA from the mussel *Mytilus galloprovincialis* and the echinoderm *Holothuria tubulosa* was also analyzed. The kinetics of *Bal*31 hydrolysis of high molecular mass DNA from the three marine invertebrates revealed a marked decrease over time of the hybridization with the cloned $(TTAGGG)_{22}$ sequence, concomitantly with a progressive shortening of the positively reacting DNA fragments. This revealed a marked susceptibility to exonuclease consistent with terminal positioning on the respective chromosomal DNAs. In full agreement, FISH results with the $(TTAGGG)_{22}$ probe showed that the repeat appears located in telomeric regions in all chromosomes of both bivalve molluscs. The presence of $(TTAGGG)_n$ repeat tracts in marine invertebrate telomeres points to its wider distribution among eukaryotic organisms and suggests an ancestry older than originally presumed from its vertebrate distinctiveness.

Introduction

The ends of eukaryotic chromosomes are capped with functional nucleoprotein structures known as telomeres, which are required to complete the telomerase-dependent replication of the tips of the linear DNA molecules and to preserve the stability and integrity of chromosome arms, as well as for chromosome positioning and segregation (1). Telomeres are structurally complex and contain several DNA components. Essentially, short double-stranded DNA repeats organized in tandem arrays at the tip regions and more

complex satellite DNA sequences attached to the tandem repeats, which constitute the internal telomere-associated DNA and form the subtelomeric regions (2). The former contain G-rich strands that are enzymatically elongated by the reverse transcriptase telomerase as single-stranded tails that extend beyond the complementary C-rich strand toward the chromosomal 3' termini (3).

The structural organization and function of telomeres are fairly conserved among widely divergent organisms from protozoa to vertebrates and higher plants (4). In contrast, telomeric DNA sequences appear to be variable between species and confined within large taxonomic groups, therefore telomeric repeats are considered group-specific (5). In this regard, vertebrates display a repeat motif, namely $(TTAGGG)_n$, which is conserved in all species so far examined from mammals to fish (6, 7). Studies on telomeric DNA in invertebrates are less abundant and mainly restricted to insects (8, 9), some other arthropods (10), and a few flat and roundworms (11–14). Invertebrate telomeric repeats differ from those found in vertebrates in several respects. They exhibit a certain degree of heterogeneity in DNA sequence and repeat lengths (1). In addition, most insects display the pentanucleotide $(T_2AG_2)_n$ as the telomeric repeat element (10). Moreover, synthetic oligonucleotides mimicking the pentanucleotide motif do not recognize sequences from vertebrate genomic DNA, nor do $(TTAGGG)_n$ oligomers hybridize with insect DNA (8). To date, studies on telomeric DNA in marine and freshwater invertebrates are rather scarce and fragmentary. For instance, $(T_2AG_2)_n$ -containing telomeres have been reported in a freshwater crustacean (10) but found absent in a holothuroid, whose genomic DNA yielded fuzzy hybridization signals with a $(TTAGGG)_n$ probe (8). A similar probe has been preliminarily reported to hybridize to some extent with genomic DNA from an echinoid (15), two marine annelids (16), a neogastropod (17), and the bay scallop (18). Some fluorescent *in situ* hybridizations to chromosomes of the pacific oyster (19) and the freshwater snail *Biwamelania habei* (20) have also been described.

We recently characterized a family of *Hin*dIII satellite DNAs in the genome of the bivalve mollusc *Donax trunculus* (21). Restriction endonuclease digestions of sperm DNA from the truncated wedgeshell with *Hin*dIII allowed detection of a DNA fragment the size of a satellite pentamer, which resisted endonuclease cleavage even under extensive digestion conditions. Cloning of the DNA in the corresponding band yielded a set of recombinants showing positive albeit weak reactivity toward one of the characterized *Hin*dIII satellites. The 836-bp cloned insert appeared to be of a heterogeneous nature, since it consisted of a satellite DNA sequence tract; preceded upstream by a segment 130 bp long made up of tandemly arrayed perfect copies of the hexanucleotide (TTAGGG). The 5'-end of the cloned insert consisted of a segment without any sequence elements resembling those of the characterized satellite structures.

Studies of telomeric repeats and their modes of association with repetitive DNA sequences might provide insight into the structural organization and function of telomeric and subtelomeric regions. In this paper we report the molecular cloning of the clustered tandem copies of the hexanucleotide (TTAGGG) previously detected in the sperm DNA of *D. trunculus*. This cloned DNA segment was used as a probe to study the chromosomal localization of $(TTAGGG)_n$ sequences in the truncated wedgeshell clam by following the time course of digestion of DNA with the exonuclease *Bal*31. We also examined genomic DNA from the Mediterranean mussel *Mytilus galloprovincialis* and the sea cucumber *Holothuria tubulosa*. The results indicate that both molluscs as well as the echinoderm species contain clusters of the (TTAGGG) repeat and that the tandem arrays are preferentially located at their chromosome ends. In addition, FISH[iii] on metaphase chromosomes of both molluscs together with RFGE analyses supported the localization suggested by the exonuclease experiments. This work demonstrates the presence and location of the vertebrate-type hexamer repeat in telomeres of marine molluscs and echinoderms.

Experimental procedures

Organisms and isolation of genomic DNA

Adult specimens of the truncated wedgeshell (*D. trunculus*), the blue mussel (*M. galloprovincialis*), and the sea cucumber (*H. tubulosa*) were either obtained from commercial suppliers or collected on the Mediterranean or northwestern coast (Ria de Ribadeo and Balcobo beach) of Spain during the breeding season. Specimens of live *H. tubulosa* were moved to the laboratory and kept in cold seawater until use. Sperm fluid was obtained as described previously (22). Briefly, mollusc shells were carefully opened with a scalpel, and the gills were removed to expose gonadal tissue. Sperm fluid was collected through a small incision with the aid of a Pasteur pipette. Male gonads from the echinoderm species were excised immediately before use, squeezed, and the resulting sperm fluid filtered through flannelette as detailed elsewhere (21). High molecular mass DNA was isolated and purified from fresh sperm suspensions by standard phenol extraction with some modifications (23).

*Bal*31 exonuclease digestion and DNA restriction

To test for the chromosomal positioning of (TTAGGG) repeat tracts, purified sperm DNA samples were subjected to *Bal*31 exonucleolytic trimming with time. High molecular mass DNA (20µg) in 20 mM Tris-HCl (pH 8.0), 600 mM NaCl, 12.5 mM $CaCl_2$, 12.5 mM $MgCl_2$, 1 mM Na_2EDTA was supplemented with λ-DNA *Hin*dIII fragments (0.5µg) to monitor the extent of the digestion and treated with *Bal*31 nuclease (2 units) at 30°C in a final volume of 150 µl. Aliquots of 3.3 µg of DNA (one-sixth of the bulk reaction) were taken at various times: time 0 (prior to enzyme addition), 10, 20, 30, 40, and 50 min, respectively. Reactions were halted by addition of Na_2EDTA to 50 mM, inactivated at 75°C for 10 min, and chilled on ice. Digested DNAs were then recovered by ethanol precipitation and finally dissolved in distilled water. A sixth part of the DNA digests (0.55 µg) was used to monitor λ-*Hin*dIII fragment trimming, whereas the remainder 2.75 µg) was further digested with *Alu*I.

Southern transfers and hybridization conditions

All enzymatic digests were electrophoresed on 0.8% agarose gels and the resolved DNA fragments subsequently transferred to positively charged nylon membranes by alkaline blotting in 0.4_N NaOH after partial depurination (24). DNA probes were labeled with fluorescein-12-dUTP by random priming with the Klenow fragment of DNA polymerase I using the Ready-To-Go labeling beads (Amersham Biosciences). Hybridizations were carried out overnight at 42°C in 50% formamide containing 0.25 M Na_2HPO_4 (pH 7.2), 7 % SDS, 1 mM EDTA, and 50 µg/ml tRNA, followed by stringent washes in 0.1 × SSC (saline-sodium citrate), 1% SDS at 65°C, except for the cloned histone H4 probe from *H. tubulosa*, which was hybridized at 35°C, and the membrane washed at 57°C. Stringency washes were followed by blocking with 0.2% casein, 0.5% SDS in phosphate-buffered saline and the filters finally reacted with an alkaline phosphatase-conjugated anti-fluorescein antibody (Tropix). Hybridization signals were visualized by chemiluminescence using the dioxetane CDP-Star (disodium 2-chloro-5-(4-methoxyspiro{1,2-dioxetane-3,2'-(5'-chloro) tricyclo[3.3.1.1^{3,7}]decan}-4-yl)-1-phenyl phosphate) (Roche Diagnostics) and recorded on x-ray film.

Dot-blot quantitations

The relative genomic abundance of the (TTAGGG) hexamer sequence was determined by dot-hybridization of graded amounts of both *D. trunculus* sperm DNA and the recombinant plasmid containing the 148-bp fragment encompassing the $(TTAGGG)_{22}$ tandem repeat. DNA samples were spot-blotted onto nylon and the membrane subsequently probed with the repetitive insert released from the recombinant clone and ^{32}P-labeled by random priming (25). After exposure to film the intensities of the radioactive signals were quantified using a computer-assisted laser densitometer loaded with the ImageQuant program (Molecular Dynamics). Mean values were derived from two independent experiments.

Rotating-field gel electrophoretic analysis

The genomic organization of (TTAGGG) tandem arrays was examined by RFGE. Aliquots of purified sperm suspensions from *D. trunculus* were embedded in 0.5% agarose plugs at a DNA concentration of 0.5 μg/μl as previously described (22). Agarose plugs containing high molecular mass DNA larger than 400 kb were incubated with selected restriction endonucleases and the resulting large genomic fragments resolved on 1.2% agarose gels in 0.5 × TBE (Tris-borate-EDTA) at 11 °C. Electrophoresis was run at 100 V for 1 h, followed by successive pulses of 10 s for 15 h and 20 s for 20 h at 200 V with 120° reorientation angles. Gels were then visualized by ethidium bromide staining and the DNA fragments transferred to a nylon membrane. Subsequent hybridizations to the cloned $(TTAGGG)_{22}$ probe were carried out as described above.

Chromosome preparation and Fluorescence *in situ* hybridization

Truncated wedgeshell and mussel specimens were continuously fed with *Isochrisis galbana* microalgae for 10 days in the laboratory. Before use, following treatment with 0.005% colchicine for 6–8 h, gills were dissected and metaphase spreads prepared as described previously (26). FISH was carried out with the $(TTAGGG)_{22}$ cloned probe labeled with digoxigenin by a standard PCR procedure and denatured at 75 °C for 15 min. The hybridization was performed in a PTC-100 microscope slide thermal cycler (MJ Research), with a solution of 50% formamide, 10% dextran sulfate in 2 × SSC, containing sonicated salmon sperm DNA (0.3 mg/ml) and the denatured digoxigenin-labeled hexamer repeat (3.3 μg/ml). Post-hybridization washes were performed in 2 × SSC at 42 °C and then sequentially with 20% formamide in 0.2 × SSC, 0.1 × SSC, and 2 × SSC for 10 min each.

For detection, slides were washed in 0.1 $_M$ Tris-HCl (pH 7.5), 0.15 $_M$ NaCl, 0.05% Tween 20 buffer and blocked in the same buffer containing 0.5% casein, but lacking detergent, at 37 °C for 30 min. The slides were then incubated with anti-digoxigenin mouse serum at a dilution of 1:200 in blocking buffer at 37 °C for 30 min, rinsed in the same buffer, and subsequently subjected to two consecutive rounds of incubation in the same conditions, first with rabbit anti-mouse serum conjugated to fluorescein isothiocyanate (FITC) at a dilution of 1:1000, and finally with FITC-conjugated goat anti-rabbit serum at the same dilution, to amplify the fluorescence signals. The slides were then washed once in blocking buffer, dehydrated through a graded ethanol series, and air dried. Metaphase preparations were counterstained with an antifade solution containing propidium iodide (50 μg/ml), examined under a microphot AFX Nikon fluorescence microscope, and photographed on Kodachrome color slide film (400 ASA).

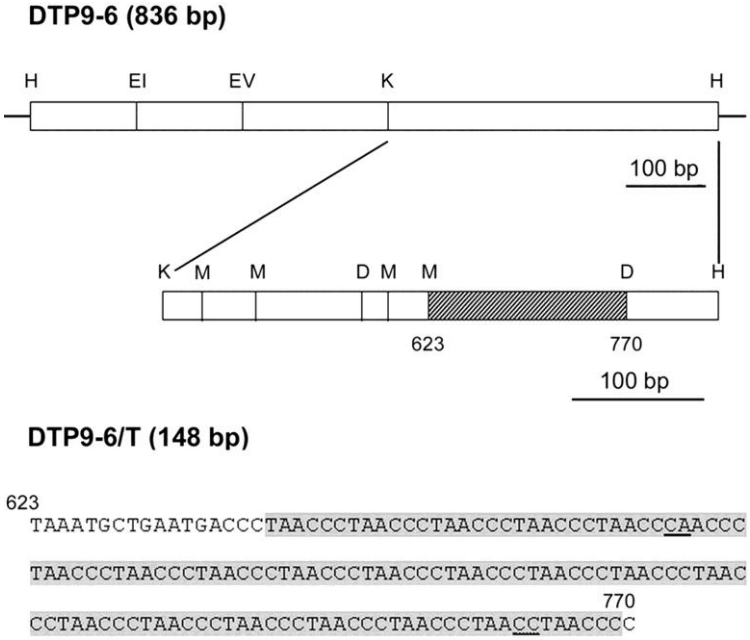
Results

Isolation and cloning of a $(TTAGGG)_{22}$ tandem repeat in *D. trunculus*

During characterization of a family of four *Hin*dIII satellite DNAs in the truncated wedgeshell, a DNA fragment of similar length to a satellite pentamer was released upon digestion of sperm DNA. This fragment resisted endonuclease fragmentation even under extensive digestion conditions. It hybridized positively in Southern blots of electrophoretically resolved restriction fragments from *Hin*dIII digests probed with the monomer clone DTHS1 (GenBank™/EBI accession number X94534) of the characterized type-1 *Hin*dIII *Donax* satellite. The DNA in the corresponding band was recovered, cloned, and sequenced (GenBank™/EBI accession number X94546) (21). The cloned insert was 836 bp long and ended in a 66-bp tract that corresponded to the 3′-terminal half of the DTHS1 monomer unit. The latter sequence was preceded upstream by a 130-bp segment comprising 22 tandemly arrayed copies of the hexanucleotide C_3TA_2, which appeared to be the reversed complement of the vertebrate-type (TTAGGG) telomeric repeat,

as well as that of the main subrepeat element found in the DTHS1 satellite DNA. The hexamer repeats were all perfect copies except for the 3′-penultimate repeat in which one C was lacking. A second internal deletion of a single T was observed 5′ contiguous to position 670 of the insert sequence. The clustered repeats were flanked on both sides by endonuclease recognition sequences for *Mse*I and *Dde*I, with cleavage sites 3′ to positions 622 and 770, respectively (Fig. 1). These two restriction enzyme sites allowed to release from the original clone a DNA fragment encompassing the entire set of highly conserved tandem repeats. The excised 148-bp DNA fragment was blunt-ended and subcloned into the *Eco*RV site of the pBluescript $^{+}$SK vector. In turn, the cloned insert was released with polylinker enzymes contiguous to the cloning site and used as a probe throughout the experimental work presented here.

Figure1. **Isolation and subcloning of the (TTAGGG)** $_{22}$ **tandem sequence from *D. trunculus* sperm DNA.** Schematic restriction map of the 836-bp genomic clone (*DTP9-6*), isolated upon complete digestion of genomic DNA with *Hin*dIII (21), is shown. The 836-bp DNA insert was cut into two fragments of 433 and 403 bp in length through the single *Kpn*I cleavage site, 3′ to position 433. The shortest restriction fragment was recovered and double digested with *Mse*I and *Dde*I to yield a 148-bp-long DNA fragment (*DTP9-6/T*) encompassing 22 tandem copies of the (GGGTTA) hexamer sequence shown *hatched*. The latter fragment was purified, blunt-end-ligated to *Eco*RV-linearized pBluescriptII^{+}SK vector, and subcloned in competent *Escherichia coli* Xl1-blue cells. *H*, *EI*, *EV*, *K*, *M*, and *D* denote *Hin*dIII, *Eco*RI, *Eco*RV, *Kpn*I, *Mse*I, and *Dde*I recognition sites, respectively. The DTP9-6 cloned sequence is available from the GenBank^{TM}/EBI databases under accession number X94546.



DNA sensitivity to exonuclease *Bal*31 and terminal location of (TTAGGG) repeat sequences in *D. trunculus*

The genomic localization of (TTAGGG) $_{n}$ sequences was examined by following the time course of *Bal*31 degradation of sperm DNA from the wedgeshell clam. High molecular mass genomic DNA extracted from fresh sperm suspensions was subjected to *Bal*31 digestion with time. Phage λ-DNA *Hin*dIII fragments were added to the DNA samples to monitor the extent of exonucleolytic trimming. DNA aliquots were taken at intervals, and a fraction of each aliquot was subsequently digested to completion with *Alu*I. DNA samples, prior and after *Alu*I digestion, were then electrophoresed in parallel on a 0.8% agarose gel (Fig. 2*A*) and Southern blotted to a nylon membrane. The half of the membrane containing the undigested *Bal*31-trimmed DNA was cut out and hybridized with fluorescein-labeled λ-DNA *Hin*dIII fragments (Fig. 2 *B*). It can be clearly observed the progressive shortening of the λ-*Hin*dIII fragment lengths. All fragment sizes gradually decreased with the course of *Bal*31 digestion as a consequence of the continuous exonuclease trimming of the λ-DNA fragment termini. Obviously, this pattern was more apparent in the smaller fragments, which tended to fade away at late time intervals.

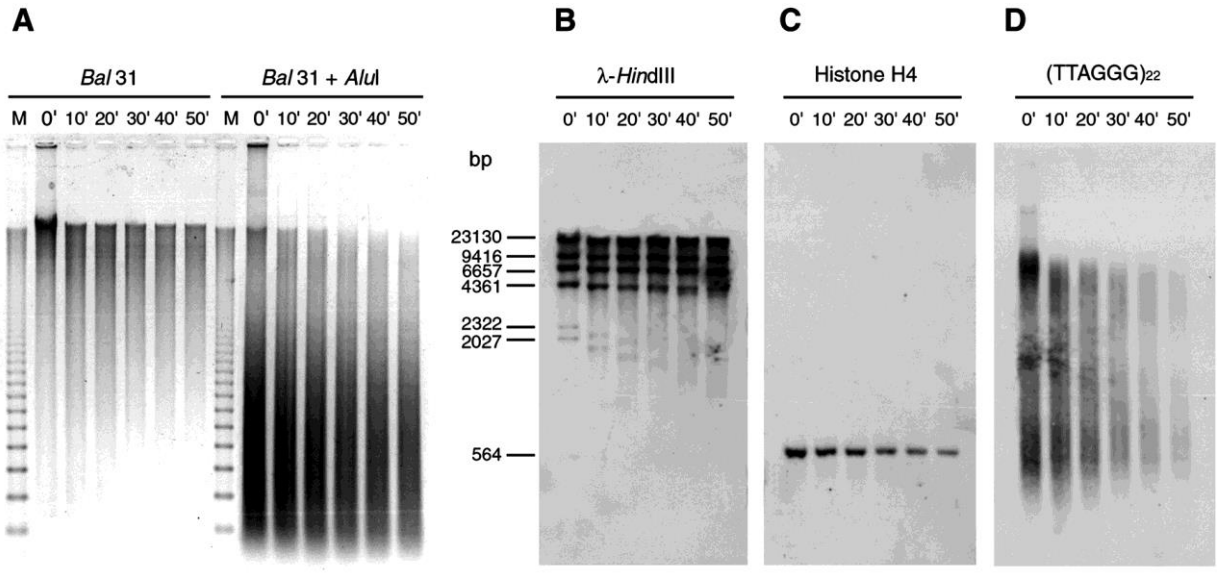


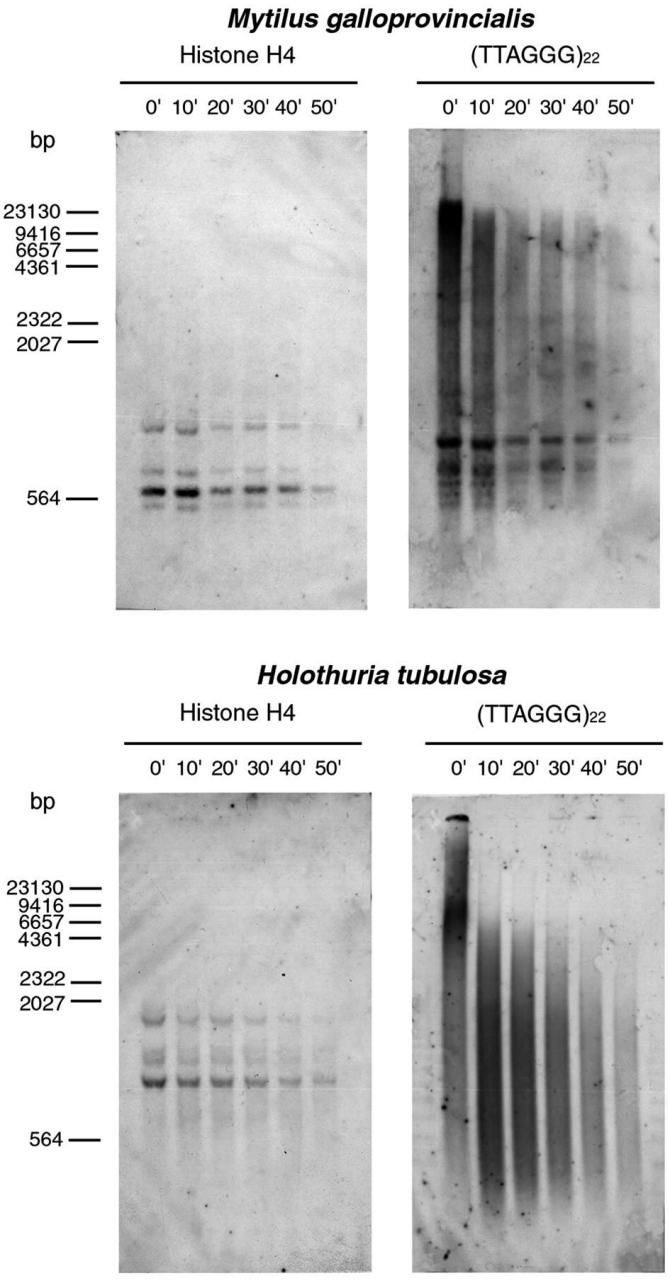
Figure2. Time course of exonuclease digestion of genomic DNA from *D. trunculus*. High molecular mass sperm DNA (20 µg) was mixed with λ-DNA *Hin*dIII restriction fragments (0.5 µg) in a final volume of 150 µl. An aliquot (25 µl) was removed prior to enzyme addition, and the remainder was digested with *Bal*31 (2 units) at different time intervals. Aliquots (25 µl each) of the bulk reaction were taken at the indicated intervals, brought to 50 mmNa_2EDTA, heat-inactivated for 10 min, and the DNA was ethanol-precipitated in the cold and dissolved in distilled water. ***A***, electrophoretic patterns of *Bal*31-trimmed DNAs (one-sixth of each digest), resolved on 0.8% agarose gels containing ethidium bromide. The remainder of the trimmed DNA was digested to completion with *Alu*I prior to electrophoresis on the same gel. ***B***, autoradiogram of the *left half* of the gel in *A* after Southern transfer and hybridization to fluorescein-labeled λ-DNA *Hin*dIII fragments. Note the gradual shortening of the λ-*Hin*dIII fragments with the time course of *Bal*31 digestion. ***C***, autoradiograph of the *right half* of the gel in *A* (*Bal*31 + *Alu*I digests) processed as in *B* but probed with a recombinant plasmid carrying the *H. tubulosa* histone H4 gene (27). Note that the histone gene appears resistant to exonucleolytic trimming. ***D***, x-ray film of the membrane in *C* after rehybridization with the cloned $(TTAGGG)_{22}$ sequence labeled with fluorescein. Note the sensitivity to *Bal*31 of digested DNA reacting positively with the $(TTAGGG)_{22}$ probe, indicating a terminal positioning of the repetitive arrays. A 142-bp DNA ladder was used as size marker (*lanes M*).

The second half of the nylon membrane, containing the *Bal*31-*Alu*I-digested DNA fragments, was initially hybridized to the insert of the recombinant pUC19 clone carrying a 1.5-kb *Hin*cII fragment containing the entire sequence of the *H. tubulosa* histone H4 gene (GenBank™/EBI accession number Z46226) (27), labeled with fluorescein (Fig. 2 *C*). As expected from the internal location of this histone gene in the genome, the DNA bands reacting with the histone H4 probe remained unaltered during the time course of *Bal*31 digestion except for a slight fall in signal intensity at longer times. The results of the hybridizations with the λ-*Hin*dIII DNA and histone H4 probes indicate the integrity of the extracted genomic DNA, ruling out any nicking or inner degradation of the DNA, while corroborating the exonucleolytic specificity of *Bal*31. Subsequently, the H4 probe was stripped off the membrane, which was rehybridized to the fluorescein-labeled $(TTAGGG)_{22}$ cloned sequence (Fig.2 *D*). In contrast to the electrophoretic patterns of the bulk DNA fragments shown in Fig. 2 *A*, where the effects of the digestions with *Bal*31 alone or in pairwise combination with *Alu*I were hardly detected by the ethidium bromide staining, the double digestion of sperm DNA generated an uneven and dauby pattern of hybridization with the repeat probe. *Alu*I digestion of genomic DNA at time 0 of the serial digestions with *Bal*31 produced a broad, intense hybridizing band of low mobility. A fast-moving band, comparable with the former but broader and less intense, was also seen near the migration front (*lane 0′* in Fig. 2 *D*). Integration of the areas under the signal peaks in densitometer tracings of the autoradiogram yielded a DNA distribution in both peaks amounting, respectively, to approximately 45 and 35% of the total DNA hybridizing to the probe in the lane. Likewise, the estimated

size of the DNA fragments under the peaks, as deduced from the scans, ranged from 10,000 to 20,000 bp and 500 to 1000 bp, respectively (data not shown). The signal intensity of the hybridizing DNA bands generated by *Alu*I digestion gradually decreased during the *Bal*31 exonucleolytic trimming, shifting toward lower length distributions with a concomitant reduction of the hybridization signals at the longest times. However, the sensitivity of the upper DNA signal to *Bal*31 was slightly higher than that of the lower signal. The intensity of the former decreased drastically within 30 min of digestion, thereafter fading faster than the lower signal, traces of which still persisted at longer digestion intervals.

The high level of hybridization of genomic DNA to the cloned $(TTAGGG)_{22}$ sequence reveals the presence of arrays of this vertebrate-type hexameric repeat within *Donax* DNA. Furthermore, the preferential susceptibility of the DNA sequences positively reacting with the repeat probe to *Bal*31 exonucleolytic trimming favors a terminal positioning of the bulk of the repeat arrays and therefore allows to qualify them as true telomeric DNA located at the chromosome ends in this mollusc.

Detection of identical hexameric repeat sequences in the mussel and sea cucumber genomes



To verify the occurrence of similar hexameric sequences in the genomes of taxonomically related molluscs and other marine invertebrates, sperm DNA samples from the blue mussel *M. galloprovincialis* and the echinoderm *H. tubulosa* were subjected to *Bal*31 digestion in the above conditions (Fig. 3). As expected, the histone H4 repeat taken as internal gene marker in the *Donax* genomic DNA showed comparable behavior in both the mussel and the sea cucumber sperm DNAs when subjected to serial digestions with *Bal*31 and to completion with *Alu*I. The corresponding DNA restriction fragments reacting positively with the histone H4 probe remained insensitive to exonucleolytic trimming. However, the hybridization patterns of the histone H4 gene were not strictly identical in the three invertebrate species examined. The minor dissimilarities observed can be attributed to varying susceptibilities of the respective DNAs to *Alu*I together with differing structural arrangements of histone H4 genes in these organisms (27). Both sperm DNAs reacted similarly to *Donax* DNA (see Fig. 2 *D*), yielding broad patterns upon digestion with *Alu*I and hybridization to the labeled $(TTAGGG)_{22}$ probe. The size distribution of both blurred patterns gradually decreased with the digestion, displaying a clear tendency to fade with time. This indicates a marked sensitivity of the hybridizing DNA to exonuclease trimming. These data support tracing the *Bal*31-sensitive DNA to chromosomal termini in both the mussel

and sea cucumber DNAs and unambiguously confirm the presence of a substantial proportion of (TTAGGG) repeats at the telomeres of these marine invertebrate DNAs.

Figure 3. Kinetics of exonuclease digestion of genomic DNA from *M. galloprovincialis* and *H. tubulosa*. To compare sperm DNAs from the blue mussel and the sea cucumber with genomic DNA from the wedgeshell clam, the former were subjected to the same serial digestions with *Bal*31 and *Alu*I enzymes as in Fig. 2. Both blots in panels to the left were probed with the plasmid containing the sea cucumber histone H4 gene (27) labeled with fluorescein. The *rightmost panels* show the same filters rehybridized with the fluorescein-labeled $(TTAGGG)_{22}$ sequence after removal of the histone probe. Note that the size of the DNA band reacting with the H4 probe is unaltered, whereas the *Bal*31-*Alu*I-digested DNA hybridizing with the cloned hexamer repeat becomes trimmed with time, indicating a terminal location.

No positively hybridizing DNA bands resistant to *Bal*31 were seen in the DNAs analyzed except those of the mussel. The hybridization pattern of *Mytilus* DNA digested with *Alu*I displayed a few discrete bands, albeit faint, in the range 500–1000 bp that were unaffected by the *Bal*31 digestion. The insensitivity to trimming of these DNA bands containing (TTAGGG) sequences suggests an internal positioning within genomic DNA, namely, at interstitial chromosome regions (2). The presence of these internal hexameric sequences implies that they are represented within genomic DNA in the mussel, constituting discrete DNA elements that produce well defined bands upon digestion. The intensity of the hybridizing bands suggests the repetitiveness of these internal sequences. Concurrently, the failure to detect similar DNA bands in the hybridization patterns from *D. trunculus* and *H. tubulosa* does not preclude their existence provided that they are either organized in DNA tracts variable in length and/or so poorly represented in the genomes that they remain undetectable.

Genomic content of (TTAGGG) repeats in the wedgeshell

To derive the content of (TTAGGG) repeated sequences at the ends of *Donax* chromosomes, the relative genomic abundance of the hexameric DNA sequence was determined from dot-blots of increasing amounts of the recombinant clone carrying the $(TTAGGG)_{22}$ insert, together with graded amounts of total sperm DNA from the mollusc. Subsequently, the nylon membrane was hybridized to a ^{32}P-labeled $(TTAGGG)_{22}$ insert released from the same clone and the hybridization signals quantified using a laser densitometer (Fig. 4). The genomic abundance computed for the hexameric tandem repeat comprised 0.05% of the total sperm DNA. Since the size of the haploid DNA complement (C-value) of the wedgeshell clam has been estimated as 1.4×10^9 bp (28), the former value roughly amounts to 4700 copies of the $(TTAGGG)_{22}$ sequence per haploid genome, equivalent to approximately 100,000 hexamer copies.

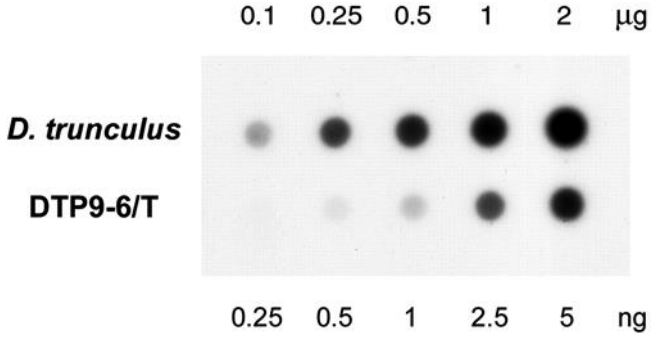


Figure 4. Relative abundance of the (TTAGGG) hexamer repeat in the *D. trunculus* genome. Graded amounts of sperm DNA (*top*) and the recombinant clone DTP9-6/T (21) carrying the $(TTAGGG)_{22}$ repeated motif (*bottom*) were dot-blotted onto a nylon membrane and hybridized to the ^{32}P-labeled $(TTAGGG)_{22}$ insert excised from the latter clone. After exposure to film, the radioactive signals were quantified using a computer-assisted laser densitometer with the ImageQuant software. Mean values were derived from two independent experiments.

The genomic length of the putative telomeric arrays of (TTAGGG) repeats can be approximated assuming their preferential localization at the ends of the mollusc chromosomes and using the repeat copy number found, together with the chromosome number of the haploid cells from *Donax* (*n* = 19) (29). Since the frequency found for the (TTAGGG) repeat amounts to 100,000 copies and taking into account the haploid chromosomal complement, the computation yields a value of 5200 repeats per chromosome or 2600 per chromosomal terminus, equivalent to a repeat tract length of 15.6 kb of DNA. This value is consistent with the size range of the upper DNA band generated by *Alu*I at zero time of the *Bal*31 digestion (*lane 0'* in Fig. 2 *D*), as well as with those from the four-cutter endonuclease digestions shown in Fig. 6, and falls within the range of lengths reported for telomeric repeats in mammalian cells (3).

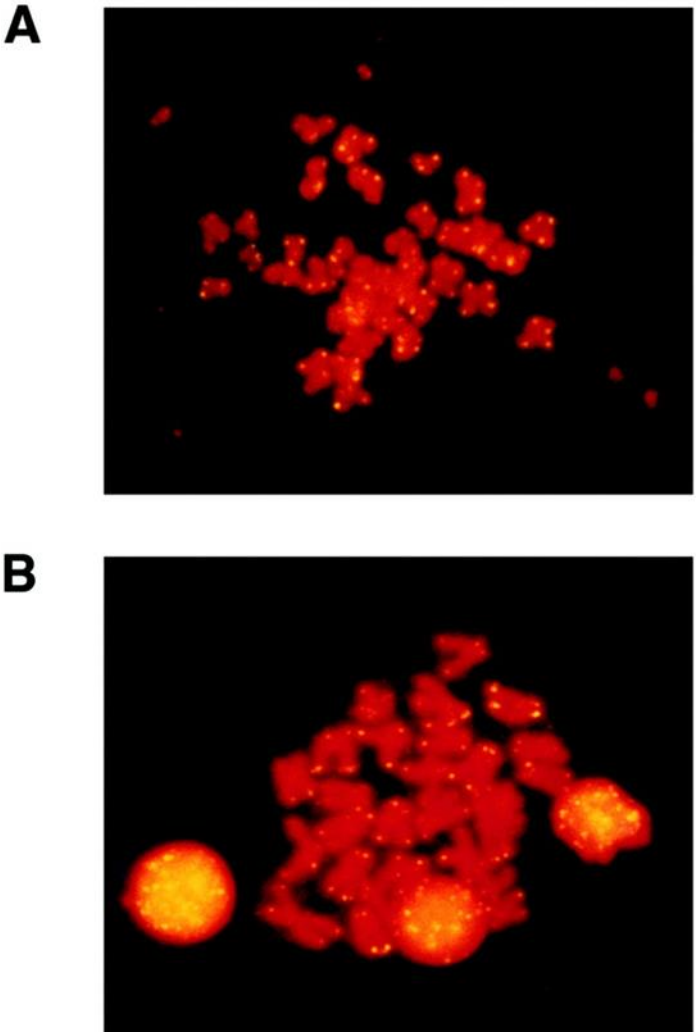


Figure 5. Chromosome localization of (TTAGGG) repeat arrays in *D. trunculus* and *M. galloprovincialis*. Metaphase chromosome squashes prepared from dissected gills of the truncated wedgeshell (A) and the common mussel (B) were hybridized to the digoxigenin-labeled $(TTAGGG)_{22}$ cloned sequence. Detection was performed by immunofluorescence with mouse anti-digoxigenin antibody and a combination of FITC-conjugated to rabbit anti-mouse and to goat anti-rabbit sera. Metaphase preparations were counterstained with antifade-containing propidium iodide. The yellow *fluorescent* signals at the ends of all chromosomes on both mollusc species, outlined against the *orange* chromosomal background, correspond to clustered (TTAGGG) repeats.

Cytogenetic localization of (TTAGGG) tandem repeats

To map the positioning of clusters of the hexameric repeat, fluorescent *in situ* hybridizations were performed on both *Donax* and *Mytilus* metaphase chromosomes. Hybridizations with the $(TTAGGG)_{22}$ cloned probe produced sharp and intense signals on all chromosomes of *Donax* (Fig.5*A*) and *Mytilus* (Fig. 5*B*). In both invertebrates, the fluorescence signals were essentially confined to the ends of all chromosomes, whereas no signals were detected in non-telomeric regions of *Donax*. In contrast, a few weak signals could be detected at internal locations of some mussel chromosomes. These hybridization signals would imply the presence of DNA sequences homologous to the $(TTAGGG)_{22}$ probe at internal loci of some chromosomes of the mussel. The existence of such internal repeated DNA tracts is consistent with the insensitivity to *Bal*31 of the discrete bands observed in the hybridization patterns of *Mytilus* DNA digested with *Alu*I as shown in Fig. 3.

The results of the cytogenetic analyses fully agree with those of the DNA sensitivity to the exonuclease *Bal*31, therefore lending further support to a preferential telomeric localization of vertebrate-type $(TTAGGG)_n$ repeats in metaphase chromosomes of the marine invertebrate species examined.

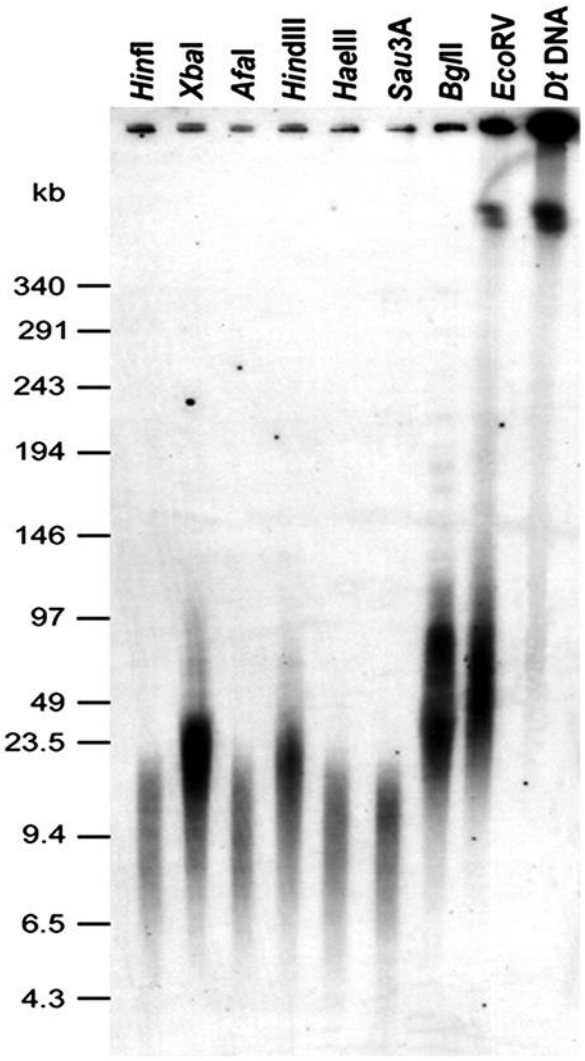


Figure 6. Genomic distributions of (TTAGGG) repeat tracts in *D. trunculus*. High molecular mass DNA from purified sperm suspensions of the truncated wedgeshell embedded in agarose plugs was digested with a selection of four-cutter (*Afa*I, *Hae*III, *Hin*fI, and *Sau*3AI) and six-cutter (*Bgl*II, *Eco*RV, *Hin*dIII, and *Xba*I) endonucleases and the resulting DNA digests subjected to RFGE analysis. Size-fractionated DNA fragments were blotted over a nylon membrane, probed with a fluorescein-labeled $(TTAGGG)_{22}$ cloned sequence, and the hybridization monitored by chemiluminescence. λ-DNA *Hin*dIII restriction fragments and a ladder of λ-phage concatemers were used as size markers. A plug containing undigested sperm DNA was loaded on the *rightmost lane* (*Dt*). Note that the undigested DNA is longer than 400 kb and that the bulk of it is retained in the application slot.

Long range organization of the (TTAGGG) repeated sequences

To examine further the genomic organization of the (TTAGGG) tandem repeats, RFGE resolution of large genomic fragments generated upon digestion of agarose-embedded high molecular mass DNA from *D. trunculus* sperm cells with various endonucleases, followed by Southern transfer and hybridization to the cloned tandem copies of the (TTAGGG) sequence, was undertaken. All blots yielded patterns of single smeary bands intensely hybridizing to the fluorescein-labeled probe, with no distinguishable internal discrete DNA fragments (Fig.6). Four-cutter endonucleases generated bands in the 9.5–15-kb size range as deduced from densitometer tracings of the autoradiograms. Restriction enzymes with six nucleotide recognition sites produced comparable banding patterns, but with larger bands, ranging from 20 kb (*lanes Xba*I and *Hin*dIII) to over 100 kb (*lane Eco*RV). *Bgl*II digestion yielded a bimodal length

distribution of DNA fragments, the lower distribution averaging 20 kb, whereas the larger centered at about 90 kb.

The tendency of all assayed endonucleases to produce single albeit broad sets of genomic DNA fragments strongly hybridizing to the $(TTAGGG)_{22}$ cloned probe under stringent conditions indicates that the hexameric repeat sequence occurs in fairly long arrays in the *Donax* genome and also points to a heterogeneity of telomere restriction fragment lengths (30). This assumption is further supported by the absence of discrete DNA fragments in the restriction enzyme digestion patterns, which suggests that hexameric sequence tracts uniform in length are not present in the *Donax* genome or, at least, so poorly represented that they elude detection. The distribution in length of the DNA fragments generated upon digestion with four-cutter enzymes is consistent with the value for the repeat length derived from the genomic content of the (TTAGGG) repeats. The higher sizes resulting from six-cutter endonucleases can be ascribed to the lower frequency of restriction enzyme sites for these enzymes in genomic DNA. Moreover, the wider size range observed may represent heterogeneous tracts of DNA consisting of hexameric repeats of comparable repetition length linked to flanking DNA sequences of variable lengths. The latter would extend to respective cleavage sites dissimilarly positioned in the telomere-associated DNA sequences at the different chromosome ends, as previously proposed for the dipteran insect *Chironomus pallidivittatus* (31).

Discussion

Here, we identified $(TTAGGG)_n$ sequence tracts in the genome of the truncated wedgeshell *D. trunculus* by Southern hybridization with a $(TTAGGG)_{22}$ probe, subcloned from an 836-bp DNA fragment isolated during the previous characterization of a *Hin*dIII satellite DNA family in the sperm DNA of the mollusc (21). The clustered tandem copies of the hexamer repeat were traced to terminal telomeric regions in all *Donax* chromosomes by exonuclease *Bal*31 sensitivity assays, and the location was corroborated by FISH analysis. Occurrence of identical (TTAGGG) sequence arrays was also confirmed in the genomes of the blue mussel *M. galloprovincialis* and the sea cucumber *H. tubulosa*.

The high levels of hybridization to the cloned $(TTAGGG)_{22}$ probe can be considered a clear indication of the existence of arrays of the vertebrate-type hexamer repeat within the genomes of the marine invertebrates surveyed. In addition, the observed sensitivity to *Bal*31 exonuclease trimming of the DNA sequences positively reacting with the repeat probe hint at a terminal localization of the bulk of the repeat arrays and, consequently, support their consideration as canonical telomeric DNA located at or close to the chromosome ends in the aforementioned invertebrate species. The latter suggestion is further sustained by the results of the FISH analyses performed, which fully agree with those of the susceptibility to the exonuclease *Bal*31. This notwithstanding, the discrete bands containing (TTAGGG) repeat tracts seen in the hybridization patterns of *Mytilus* DNA digested with *Alu*I, unaffected by *Bal*31, imply that they constitute repetitive DNA stretches of comparable lengths, interstitially located in the chromosomes of the blue mussel. In this concern, it is worth mentioning that telomeric repeats have been found at internal sites on the chromosomes of many organisms (2, 6).

Although telomeric DNA in most organisms consists of arrays of short repeats, there is no universal telomere motif in eukaryotes, but rather repeat sequences vary between groups of species (1). Thus, the (TTAGGG) motif has been considered to be typical of vertebrates even though it has been also observed in certain molds, fungi, and a protozoan. Despite this sequence multiplicity and interspecies differences, telomere function is rather conserved throughout the eukaryotes (5). The evidence provided here as regards the telomeric localization of tandemly arrayed (TTAGGG) repeats in *D. trunculus and M. galloprovincialis* as well as in the echinoderm *H. tubulosa* suggests a number of evolutionary considerations. Thus, the detection of the genuine vertebrate telomere motif in invertebrates argues for its wider representation among very divergent

species and, consequently, an ancestry older than initially presumed from its vertebrate distinctiveness may be surmised, aside from an equivalent function in nonvertebrate telomeres.

Moreover, the maintenance of telomeric repeats requires a functional telomerase (3). If the very early eukaryotic origin postulated for this reverse transcriptase (32) is taken into account, it may be inferred that emergence of novel telomere repeat sequences ought to occur rarely in evolution as argued in the case of the $(TTAGG)_n$ telomere sequences in insects (10). Consequentially, new and non-related appearances over evolutionary time of the vertebrate-type (TTAGGG) element in very distant organisms should be posited as improbable. Despite the heterogeneity of telomeric DNA sequences encountered throughout the eukaryotes, all bear a relative similarity and retain equivalent functional and structural features. Although the (TTAGGG) hexamer cannot be unambiguously identified as a true consensus element, it is the only telomeric repeat widely represented among very divergent organisms from protozoa to mammals and invertebrates as those reported here (1). In addition, the reverse transcriptase function of telomerase favors the likelihood of the existence of an ancestral telomeric motif. The enzymatic activity of telomerase is analogous to that of the major non-long terminal repeat retrotransposon L1 of mammals, and both share substantial sequence similarities, which underscore their evolutionary relatedness and lend support to the assumption that eukaryotic cells could have recruited the retrotransposon to acquire telomerase activity at the dawn of eukaryotes (33). This potential situation could explain the similarities mentioned among eukaryotic telomere DNAs, whereas more complex and distinct telomeric ends, such as those found in most fungi (1), or the transposons at the *Drosophila* chromosome termini that elongate by transposition of telomere-specific retrotransposable elements (34), could have appeared more recently in evolution.

Telomere repeat sequences in vertebrate organisms appear to be of variable length, from short telomeres averaging 10 kb in humans (35) to 20–150 kb $(TTAGGG)_n$ sequences in mice (36). The 15.6-kb average length reported here for the tandemly repeat arrays in *D. trunculus*, estimated from Southern analysis of its genomic DNA digested with four-cutter endonucleases and confirmed by dot-blot quantitation of its relative genomic abundance and the RFGE analysis of digested high molecular mass DNA, is consistent with the range of the mentioned lengths. However, these measurements should be considered average approximations, since they do not take into account the heterogeneity of telomere restriction fragments (37) and fail to discern the internal subtelomeric repetitive DNA sequences adjacent to the terminal telomere repeats, which often appear interspersed with the latter giving rise to satellite DNAs containing degenerate variant telomeric repeats (5, 6, 38). In this regard, one member of *the Hin*dIII satellite DNA family characterized in the *D. trunculus* genome (monomer clone DTHS1) displays the (TTAGGG) hexamer as a prominent subrepeat structure, whereas the remainder contain short sequence motifs directly related to the hexameric subrepeat (21). In addition, the 836-bp fragment from which the $(TTAGGG)_{22}$ probe used here was subcloned ended with a DNA stretch encompassing the 3′-terminal half of the DTHS1 monomer unit. The failure to detect *Bal*31-resistant DNA bands in the hybridization patterns from *D. trunculus* (Fig. 2) and *H. tubulosa* (Fig. 3) suggests that the tandemly arrayed (TTAGGG) repeats in these invertebrates are preferentially located at the ends of their chromosomes. Any internal (TTAGGG) tracts might be either variable in length, poorly represented in the genomes, or interspersed with satellite sequences in a degenerate manner, such as the subrepeat structures in the *Donax Hin*dIII satellite DNA family, so that they elude detection. In contrast, the resistance to *Bal*31 trimming of the few bands seen in the hybridization pattern of *Mytilus* DNA digested with *Alu*I, reacting positively with the $(TTAGGG)_{22}$ probe (Fig. 3), suggests an internal positioning of the corresponding DNA fragments in the blue mussel genome, constituting short arrayed telomeric sequences of comparable length incorporated in a dispersed manner into internal genomic locations likely containing repetitive DNA sequences.

The long range organization of the (TTAGGG) repeated sequences revealed by the RFGE analysis of digested genomic DNA from the wedgeshell clam points to the lack of internal hexamer repeat tracts uniform in size, thereby supporting further the inference that the vertebrate-type $(TTAGGG)_n$ repeat arrays

characterized in the *D. trunculus* genome are primarily located at the termini of all chromosomes. Likewise, the failure to detect these hexameric repeats in interstitial chromosomal loci suggests that the repeated sequence does not appear dispersed within the genome in this marine invertebrate to a significant extent, contrasting with many other eukaryotic genomes (39). To ascertain the arrangement of telomere-associated repetitive DNA and the interspersion pattern with telomeric sequences in the *Donax* genome and its potential involvement in telomere maintenance, more experimental data are required. To address these questions we are extending restriction analyses and cloning of DNA fragments encompassing the boundaries between flanking subtelomeric repetitive DNA and the terminal $(TTAGGG)_n$ telomere sequences described here in the truncated wedgeshell genome.

Acknowledgment

We thank Robin Rycroft for his careful reading of the entire manuscript.

Footnotes

This work was supported in part by grants (to L. C.) from the Spanish Dirección General de Enseñanza Superior (PB97-1136) and the catalan Direcció General de Recerca (SGR01-354) within the framework of the Centre de Referència en Biotecnologia de la Generalitat de Catalunya.The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Miroslav Plohl was recipient of an EMBO short term fellowship.

References

1. Zakian V. A. *(1995) Science 270:*1601–1607.
2. Biessman H., Mason J. M. (1994) *Chromosoma (Berl.) 103:*154–161.
3. Collins K. (2000) *Curr. Opin. Cell Biol. 12:*378–383.
4. Blackburn E. H. (1994) *Cell 77:*621–623.
5. Blackburn E. H. (2001) *Cell 106:*661–673.
6. Garrido-Ramos M. A., de la Herrán R., Rui Zrejón C., Rui Zrejón M. (1998) *Cytogenet. Cell Genet. 83*:3–9.
7. Meyne J., Ratliff R. L., Moyzis R. K. (1989) *Proc. Natl. Acad. Sci. U. S. A. 86*:7049–7053.
8. Okazaki S., Tsuchida K., Maekawa H., Ishikawa H., Fujiwara H. (1993) *Mol. Cell. Biol. 13:*1424–1432.
9. Saiga H., Edstrom J. E. (1985) *EMBO J. 4:*799–804.
10. Sahara K., Marec F., Traut W. (1999) *Chromosome Res. 7*:449–460.
11. Chiurillo M. A., Cano I., da Silveira J. F., Ramírez J. L. (1999) *Mol. Biochem. Parasitol. 100*:173–183.
12. Joffee B. I., Solovei I. V., Macgregor H. C. (1998) *Chromosoma 107*:173–183.
13. Muller F., Wicky C., Spicher A., Tobler H. (1991) *Cell 67*:815–822.
14. Teschke C., Solleder G., Moritz K. B. (1991) *Nucleic Acids Res. 19*:2677–2684.
15. Lejnine S., Makarov V. L., Langmore J. P. (1995) *Proc. Natl. Acad. Sci. U. S. A. 92*:2393–2397.

16. Jha A. N., Dominguez I., Balajee A. S., Hutchinson T. H., Dixon D. R., Natarajan A. T. (1995) *Chromosome Res. 3:*507–508.
17. Vitturi R., Colomba M. S., Gianguzza P., Pirrone A. M. (2000) *Genetica 108*:253–257.
18. Estabrooks S. L. (1999) *J. Shellfish Res. 18*:401–404.
19. Guo X., Allen S. K. (1997) *J. Shellfish Res. 16*:87–89.
20. Nomoto Y., Hirai M., Ueshima R. (2001) *Zool. Sci. 18*:417–422.
21. Plohl M., Cornudella L. (1997) *J. Mol. Evol. 44*:189–198.
22. Drabent B., Kim J. S., Albig W., Prats E., Cornudella L., Doenecke D. (1999) *J. Mol. Evol. 49*:645–655.
23. Sainz J., Cornudella L. (1990) *Nucleic Acids Res. 18*:85–890.
24. Reed K. C., Mann D. A. (1985) *Nucleic Acids Res. 13*:7207–7221.
25. Feinberg A. P., Vogelstein B. (1983) *Anal. Biochem. 132*:6–13.
26. González-Tizón A., Martínez-Lage A., Rego I., Ausio J., Méndez J. (2000) *Genome 43*:1–8.
27. Drabent B., Louroutziatis A., Prats E., Cornudella L., Doenecke D. (1995) *DNA Sequence 6*:41–45.
28. Hinegardner R. (1974) *Comp. Biochem. Physiol. 47A*:447–460.
29. Cornet M., Soulard C. (1990) *Genetica 82*:93–97.
30. Martens U. M., Brass V., Engelhardt M., Glaser S., Waller C. F., Lange W., Schmoor C., Poon S. S. S., Lansdorp P. M. (2000) *Biochem. Soc. Trans. 28*:245–250.
31. Zhang Y. J., Kamnert I., López C. C., Cohn M., Edström J. E. (1994) *Mol. Cell. Biol. 14*:8028–8036.
32. Malik H. S., Burke W. D., Eickbush T. H. (1999) *Mol. Biol. Evol. 16*:793–805.
33. Kazazian H. H., Jr. (2000) *Science 289*:1152–1153.
34. Biessman H., Zurovcova M., Yao J. G., Lozovskaya E., Walter M. F. (2000) *Chromosoma 109*:372–380.
35. Moyzis R. K., Buckingham J. M., Cram L. S., Dani M., Deaven L. L., Jones M. D., Meyne J., Ratliff R. L., Wu J.-R. (1998) *Proc. Natl. Acad. Sci. U. S. A. 85*:6622–6626.
36. Kipling D., Cooke H. J. (1990) *Nature 347*:400–402.
37. Zijlmans J. M., Martens U. M., Poon S. S. S., Raap A. K., Tanke H. J., Ward R. K., Lansdorp P. M. (1997) *Proc. Natl. Acad. Sci. U. S. A. 94*:7423–7428.
38. Vermeesch J. R., De, Meurichy W., Van den Berghe H., Marynen P., Petit P. (1996) *Cytogenet. Cell Genet. 72*:310–315.
39. Meyne J., Baker R. J., Hobart H. H., Hsu T. C., Ryder O. A., Ward O. G., Wiley J. E., Wurster-Hill D. H., Yates T. L., Moyzis R. K. (1990) *Chromosoma 99*:3–10.

[*] This is a post-peer-review, pre-copyedit version of an article published in [Journal of Biological Chemistry]. The final authenticated version is available online at: [doi: 10.1074/jbc.M201032200].
[†] lcmbmc@cid.csic.es
[iii] The abbreviations used are: FISH, fluorescence in situ hybridization; RFGE, rotating-field gel electrophoresis; FITC, fluorescein isothiocyanate.