The highly variable pentameric repeats of the AT-rich germline limited DNA in *Parascaris univalens* are the telomeric repeats of somatic chromosomes

Claudia Teschke⁺, Gabriele Solleder and Karl B. Moritz*

Zoologisches Institut der Universität München, Luisenstraße 14, D-8000 München 2, FRG

Received February 15, 1991; Revised and Accepted April 19, 1991

EMBL accession nos X57501–X57511, X58635

ABSTRACT

We have characterized the organization of the germline limited DNA of *P. univalens* by means of sequence analysis. The repeat unit of this satellite DNA is the pentanucleotide 5'TTGCA, although there is a high degree of sequence variation. Repeat variants are not arranged in tandem but in a disperse, nonrandom manner. In the somatic genome which arises from the germline genome through extensive genomic rearrangement early in development, copies of these pentamers represent the telomeric repeats, indicated by their sensitivity to Bal 31 and their presence in a somatic endlibrary. Unlike telomeric sequences from other species the *P. univalens* telomeres do not display consecutive guanines and no strand bias for that base, recently suggested as universal features of eukaryotic telomeres. Investigation of fragments that carry pentameric repeats along with sequences of different type identifies a 5 bp consensus sequence at the junction point. We suggest a model in which pentameric repeats originate via amplification by a terminal transferase (telomerase) in both the germline and the somatic genome.

INTRODUCTION

Genomic rearrangement is an intrinsic part of early development in the nematode genus *Ascaris*. Cytological studies show that considerable amounts of the genome are eliminated in all somatic founder cells (1). This process, called chromatin diminution, is strongly dependent on somatic quality of the cytoplasm (2, 3) and in turn is a clear marker for somatic cells. The exact significance of chromatin diminution is not known. As in *Cyclops* (4), chromatin diminution in *Ascaris* mainly discards constitutive heterochromatin.

In *Ascaris suum* and *Parascaris univalens* (= *Ascaris megalocephala*) the amount of eliminated chromatin (22% and 85%, respectively) is similar to the amount of a satellite DNA present only in the germline genome and is also similar to the size of a fast reassociating fraction also not present in the somatic

genome (5, 6). These measurements are precise enough to show that not less than 99% of the eliminated, germline limited DNA is composed of large blocks of highly repeated sequences.

For *A. suum* the organization of the germline limited satellite DNA was defined by restriction enzyme analysis (7). This DNA is composed of a 123 bp repeat unit; sequence variants occur in tandem clusters. Sequence analysis of genomic repeat monomers shows length homogeneity of the repeat unit (8), whereas sequence analysis of cloned repeat units revealed length heterogeneity (9). In *P. univalens*, restriction enzyme analysis revealed an irregular distribution of recognition sequences, for example of Sau 3A and Hinf I (10).

In *P. univalens*, chromatin diminution not only involves loss of the large terminal heterochromatic blocks but also chromosome fragmentation (1, 11). The haploid single plurivalent ('Sammel'-) chromosome disintegrates into about 30 somatic chromosomes (our unpublished results). In brief, metaphase chromosomes destined to undergo diminution show differential condensation. The terminal heterochromatic blocks flank an intercalary segment which appears in a 'beads on a string' arrangement. Condensed arrays representing the presumptive somatic chromosomes are connected by noncondensed chromatin, referred to as linker chromatin (3). Fragmentation is a consequence of the excision of that interstitial chromatin which comprises not more than 0.05% of the eliminated material.

The formation of new somatic chromosomes in *P. univalens* through diminution requires the formation of new telomeric sequences to guarantee stable chromosome ends. Until recently, diminution models basically proposed that the telomeric repeats are preexisting as silent telomeres in the unrearranged chromosomes (12, 13). There is now strong evidence that in ciliated protozoans, which show extensive genomic rearrangement during differentiation of a new macronucleus, new telomeres are formed in a telomerase mediated process (14, 15, 16). Telomerases are ribonucleoproteins that add telomeric DNA onto G-rich stranded telomeric sequences. They are specialized reverse transcriptases using an internal RNA-template (17, 18, 19).

We report that the germline limited satellite DNA of *P. univalens* is organized in a fairly unusual manner. Sequence

* To whom correspondence should be addressed

⁺ Present address: MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

variants of a pentameric repeat unit show a hyperdisperse (nonrandom) distribution pattern. We present arguments that this pattern is well explained by a non-DNA-templated elongation via a telomerase. We are aware that sequence heterogeneity is not characteristic for the telomeric repeats added by the telomerase activities found in *Tetrahymena*, *Oxytricha*, *Euplotes* and HeLa cell extracts (14, 16, 20, 21). We propose that this results from novel properties of the predicted *P. univalens* enzyme.

MATERIALS AND METHODS

Preparation of DNA from *P. univalens* tissue

Germline DNA was isolated from male or female gonads essentially as described by Roth and Moritz (7). Somatic DNA

was extracted from the 4000-g nuclear pellet of intestinal homogenates. The intestines were prepared from adult females, they are free of any other tissue, especially of gonads. After buoyant density centrifugation CsCl was removed by dialysis against TE (10 mM Tris/HCl, 1 mM EDTA, pH 8.0), and the Hoechst dye was removed by phenol extraction.

Digestions with Bal 31 and restriction enzymes

Somatic DNA was digested with Bal 31 from Bethesda Research Laboratories in 12 mM CaCl₂, 24 mM MgCl₂, 0.2 M NaCl, 20 mM Tris/HCl, 1 mM EDTA, pH 8.0 using 0.35 U enzyme/100 µg DNA at 30°C. Reactions were stopped by adding EGTA to 20 mM. Restriction enzyme digestions were carried out according to the manufacture's instructions.

```

pAm1
TTGCA TTGCA TTGCA TTGCA TTGAA TTGGA TTGAA TTGCA TTGCA TTGCA CCGCA TTGAA TTGCA TTGCA TTGTA TTGCG
TTGAA TTGCA TTGCA TTGAA TTGCA TTGAA TTGCA TTGAA TTGCA TTGCA TTGCA TTGAA TTGCA TTGCA TTGTA TTGCA
TTGCA TTGAA TTGCA TTGCA TTCAA TTGCA CTGCA TTGCA TTGAA TTGCA TTGAA TTGCA TTGCA TTGCA ATGAA TTGCA
TTGAA TTGCA TTGAA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTCAA TTGCA CTGCA TTGCA
TTGAA TTGCA TTGAA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGAA TTGCA TTGCA TTGAA TTGCA TTGCA TTGCA
TTGCA TTGCA CTGCA TTGCA TTGCA TTGCA TTGAA CTGCA TTGCA TTGCA TTGCA TTGCA TTGAA 470

pAm2
TTGCA TTGCA TTGCA TTGAA TTGCA TTGAA TTGCA TTGCA TTGCG TTGCA TTGCA TTGCG TTGAA TTGCA TTGAA TTGCA
TTTCA TTGCA TTGCA TTCAA TCTTT TTGTA TCGCA TTGAA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA
TTGCA TTGCA TTCAA TTGCA CTGCA TTGCG TTGCA TTGCA TTGCA TTGAA TTGCA TTGCA 220

pAm3
TTGCA TTGAA TTCAA TTGCA TTGCA TTGAA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TCGCA TTGAA
TTGCA TTGCA TTGCA TTGCA TTGAA TTGCA TTGCA TTGCA TTGTA TTGCA TTGAA TTGAA TTGCA TTGCA TTGCA
TTGCA TTGAA TTGCA TTGCA TTGAA TTGCA TTGCA TTGCG TTGCA CTGTA TTGCA TTGCG TTGCA CTGTA TTGCA
TTGCA TTGCA TTGTA TTGCA TTGCA TTGGA TTGCA TTGCA TTGCA TTGCA TTGCA TTCAA TTGCA TTGCA TTGCA TTGCA
TTGAA TTGCA TTGAA CTGCA TTGCA TTGCA TTGAA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTCTA TCTCT TTGTA
TCGCA TTGGA 410

pAm4
TCTCT TTGTA ACGCA TTGGA TTCAA TTGCA TTGAA TTGCA TTGCA TTGCA TTGCA TTGCA AGGAA TTGCA TTGAA TTGCA
TTGCG TTCAA TTGCA TTGCA TTGCA TTGCA TTGAA TTTCA ATACA TTGCA TTGAA TTACA TTGCA TTGCA TTGAA TTGCA
TTGCA TTGCA TTGCA TTGCA TTGAA TTGCA TTTCA TTGCA TTGAA TTGCA TTGCA TTGAA TTGCA TTGCA CTGCA
TTGCA TTGCA TTGCA TTGAA TTGCA TTGCA TTGCA TTGCA CTGCA TTGCA TTGCA TTGCA TTGAA TTGCA TTGCA
TTGCA TTGCA TTGAA TTGGA 340

pAm6
TTGCA TTGAA TTGTA TTGCA TTGGA TTGCA TTGCA TTGAA TTGCA TTGCA TTGCG TTGCG TTGCA TTGAA TTGCA TTGAA
TTGCA TTGCA TTGCA CTGTA TTGCA TTGCA TTGCA TTGTA TTGCA TTGAA TTGCA TTGAA TTGCA CTGCA TTGCA TTGAA
TTGCA TTGAA TTGCA CTGCA CTGGA TTGAA TTGCG TTGCA TTTCA TTGCA TTGAA CTGCA TTGAA TTTG TTGCA TTGCA
TTCAA TCTCT TTGTA TCGCA TTGGA 265

pAm7
TTGCA TTGAA TTGCA TTGCA TTGCA TTGAA TTGCA TTGTA TTTCC TTGAA TTGCA TTTCA TTGCA TCGCA TTGGA TTGCA
TTGAG TTGCA TTGCA TTGAA TTGCA TTGAA TTGCA TTGCA TTGCA TTGCA TTGCA TTGAA TTGCA CTGAA TTGCA
TTGCA TTGCG TTGAA TTGCA TTGAA CTGCA TTGAA TTGCA TCTCA TTGCA TTGAA TTGCA TTGAA TTGCA TTGCA TTGCA
TTGAA TTGCA TTGCA TTGCA TCGCA TAGCA TCGCA TTGCA TTGCA TTGCA TTGCA TTGAA ACGCA TTGCA TTGCT
TTGCA TCTCA TTGCA TCGCA 340

pAm8
TTGCA TTGAA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA CTGTA TTGCA TTGCA TTGCA TTGTA TTGCA GTCAA TTGCA
TTGCA TTGCA TTGAA TTGCA TTGCA TTCAA TTGCA TTCAA ATGAA CTACA TTGCA TTGCA TTGAA TTGCA TCGCA TTTCT
TTGCT TTGCA TTGCA TTGAA TTTG TTGCA TTGCA TTGCA TTGCA TTGCA TTCAA TCTCT TTGTA TCGCA TTGGA 230

pAm9
TTGCA TTGCA TTGAA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGAA TTGCA TTGAA TTGCA TTGCA TTGCA
TTGCG CTGAA CTGCA TTGAA TTGCA TTTCA TTGCA TTTCA TTGCA TTGCA TTGCA TTGCA TCGCA TTGCA TTGAA TTGCA TTGAA
TTGCA TTTCA TTGCA TTGCA TTGCA TTGCA TTGCG TTGAA TTGCA 205

pAm10
TTGCA TTGCA TTGCA TTGAA TTGCA TTGAA TTCAA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCG TTGAA TTGCA TTGAA TTGCA
TTTCA TTGCA TTGCA TTCAA TCTTT TTGTA TCGCA TTGAA TTGCA TTGCA TTGAA TTGCA TTGCA TTGCA TTGCA TTGCA
TTGCA TTGCA TTCAA TTGCA CTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGAA TTGCA TTGCA TTGAA TTGCA
TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA
TTGCA TTGCG TTGAA TTGCA TTGAA TTGCA TTTCA TTGCA TTGCA TTGCA TCTTT TTGCA TCGCA TTGGA TTGCA TTGCA
TTGAA TTGCA TTGCA TTGCA TTGCA TTGCT TTGTA TTGCA TTCAA TTGCA CTGCA TTGCA TTGCA TTGCA 470

```

Figure 1. DNA sequences of satellite DNA clones pAm1,...,pAm10. The repeat unit is displayed in pentameric blocks. To stress the disperse distribution pattern of repeat variants the prevalent repeat (Leitmotiv 5' TTGCA) is set boldface and the main variant (5' TTGAA) in italics. Only the Leitmotiv occurs in small tandem clusters. All restriction sites GAATTC (EcoR I), GATC (Hinf I), ATGCAT (Nsi I) found within the pentameric sequences are underlined. N nucleotide not determined.

Recombinant DNA

Plasmid DNA was digested with restriction enzymes as indicated in the text and dephosphorylated as described by Maniatis et al. (22). Ligations were done using 100 ng plasmid DNA and insert DNA in 3 molar excess. Transformed cells were grown in TB medium (23) and plasmid DNA isolated according to Birnboim and Doly (24) with the exception that after RNase incubation DNA was PEG precipitated (25).

DNA sequencing with Sequenase T^M

2–3 µg plasmid DNA was denatured in 0.2 N NaOH for 10 min and precipitated with 1/10 vol 2 M NH₄Ac pH 4.5, 2 vol EtOH. Sequencing was done with Sequenase T^M according to the manufacturer's (USB) instructions. Sequencing gels were run on an LKB macromould apparatus following the LKB protocol.

Probe labelling with biotin-11-dUTP

Genomic DNA and full length plasmid DNA was labelled using the nick translation kit from Bethesda Research Laboratories and Biotin-11-dUTP from Enzo. Isolated insert DNA was labelled by random hexamer priming according to Feinberg and Vogelstein (26) using 2.5 µl 0.2 mM Biotin-11-dUTP/25 µl reaction volume.

Southern blotting, colony fixation and hybridization

After size fractionation by agarose gel electrophoresis, DNA was transferred onto nitrocellulose as described by Southern (27). Recombinant colonies were fixed on Whatman 541 filters as described by Taub and Thompson (28). All hybridizations were performed with biotinylated probes in 45% formamide, 4×SSC, 20 mM NaPO₄, 10% dextran sulfate, 0.8×Denhardt's solution and 200 µg/ml denatured salmon sperm DNA at 42°C over night. Post-hybridization washes were twice in 2×SSC, 0.1% SDS at room temperature (rt) for 5 min, twice in 0.2×SSC, 0.1% SDS at rt for 5 min, once in 0.16×SSC, 0.1% SDS at 60°C for 15 min, and finally filters were rinsed in 2×SSC at rt.

Detection of biotinylated probes

Biotinylated probes were detected by the alkaline phosphatase reaction using the Detek-1-alk-complex from Enzo and following the Enzo protocol.

Construction of the somatic endlibraries

Construction was done essentially as described by Richards and Ausubel (29). 2 µg of high molecular weight somatic DNA was digested with 0.05 U Bal 31 for 30 sec at 30°C in a 200 µl reaction volume, extracted with phenol/chloroform and precipitated with EtOH. DNA was ligated to 100 ng Hinc II/BamH I double digested pUC 19 DNA in a 25 µl volume at 16°C for 22 hr using 2.5 U T4-ligase from Bethesda Research Laboratories according to the manufacturer's instructions. After phenol/chloroform extraction and EtOH precipitation DNA was either digested with BamH I and Mbo I (A) or Eco RI (B), again purified and self-ligated in 1 ml at 4°C for 22 hr using 1.2 U T4-ligase. Prior to transformation of DH5α cells the ligation mixture was heated to 65°C for 10 min. The endlibraries were screened with satellite DNA.

RESULTS

Sequence analysis of cloned germline limited satellite DNA fragments

To characterize the organization of the germline limited satellite DNA respective DNA fragments were cloned and sequenced. Germline limited satellite DNA was derived from Hoechst 33258-CsCl density gradients. For want of a defined restriction fragment (10) Sau 3A fragments of varying size were shotgun cloned. To reduce cloning instability only small fragments ranging from 100 bp to 500 bp were selected. Previous experiments had revealed a strong tendency of the satellite sequences to become deleted during propagation in *E. coli*. This behavior was much the same as observed in cloning simple sequence satellite DNA of *D. melanogaster* (30), and suggests that the repeat unit of

Table 1. Frequencies of pentameric variants

Variant	DNA source													Σ
	pAm1	pAm2	pAm3	CsCl fractionated satellite			pAm8	pAm9	pAm10	Somatic		Germline		
				pAm4	pAm6	pAm7				pTem1	pTem5	pÜ12	pÜ27	
Leitmotiv														
V ₀ TTGCA	59	24	52	39	24	36	26	25	61	132	69	158	188	893
Main var.														
V ₁ TTGaA	19	6	13	11	11	14	4	7	12	40	20	50	34	241
Minor var.														
V ₂ cTGCA	3	1	1	2	3	1		1	2	11	1	11	15	52
V ₃ TTtCA	2	2		2	1	1		3	2	10	6	6	10	45
V ₄ TTGCg	1	3	2	1	3	1		2	2	5	6	10	5	41
V ₅ TTGgA	2	2	2	2	2	1	1		2	3	1	7	6	31
V ₆ TTcCA		2	2	2	1		3		4		2	7	7	30
V ₇ TcGCA	1	1	2	1	1	4	2	1	2	3		7	3	28
V ₈ TTGtA	2	1	3	1	3	1	2		1			8	4	26
V ₉ TTcaA	2	1							2		1		11	17
V ₁₀ cTGtA			2		1		1			1		6		11
V ₁₁ -V ₄₄	3	1	3	6	3	9	7	1	3	6		20	16	78

All clones were picked by screening the respective library with satellite DNA. For origin of clones pAm1,...,pAm10, pTem1 and pTem5 see text. Clone pÜ12 was found in a germline library. Total germline DNA was partially digested with Sau 3A and ligated to BamH I cleaved pUC8-DNA. Clone pÜ27 was found in an alternate germline library. Total germline DNA was digested with Alu I, to which satellite DNA is resistant (10), the uncleaved DNA fraction was recovered by differential PEG precipitation (4%), digested with Sau 3A and integrated into the BamH I site of pUC8-DNA. Base substitutions with respect to the Leitmotiv are set in small letters.

P. univalens is also a simple sequence. In this case 100 bp to 500 bp inserts should be long enough to include a tandem repeat array.

Size fractionated Sau 3A-digested satellite DNA was cloned into pBR 322; positive clones were identified by colony hybridization using satellite DNA as a probe. Out of 10 positive clones (pAm1,...,pAm10) 9 could be stably propagated, and their inserts were sequenced.

As shown in figure 1 all satellite DNA inserts belong to the same type of simple sequence. The repeat unit is a pentamer;

Table 2. Proportions (%) of pentameric sequence variants

Clone	Leitmotiv	Main Variant	Minor Variants	Pentamers Σ
pAm1	62.8	20.2	17.0	94
pAm2	54.6	13.6	31.8	44
pAm3	63.4	15.9	20.7	82
pAm4	58.2	16.4	25.4	67
pAm6	45.3	20.8	33.9	53
pAm7	52.9	20.6	26.5	68
pAm8	56.5	8.7	34.8	46
pAm9	62.5	17.5	20.0	40
pAm10	65.6	12.9	21.5	93
pTem1	62.6	19.0	18.4	211
pTem5	65.1	18.9	16.0	106
pÜ12	54.5	17.2	28.3	290
pÜ27	62.9	11.4	25.7	299
Total	59.8	16.2	24.0	1493

Variants are not enriched in any single DNA insert. $P \approx 0.05$, in the case of pÜ27 $P \approx 0.01$ by a χ^2 test. However, V_9 and V_{10} are overrepresented in pÜ27 and pÜ12, respectively (see table 1).

Table 3. Substitutions within pentamers

Position	T	T	G	C	A
Rate of subst.	0.059	0.036	0.088	0.239	0.049
Proportion of all subst. (%)	12.5	7.5	18.8	50.8	10.4
	C 83.0	C 90.5	A 54.6	G 77.1	G 68.5
new base	A 13.6	G 5.7	T 38.6	T 13.7	T 28.8
(%)	G 3.4	A 3.8	G 6.8	C 9.2	C 2.7

Substitution rates which differ by more than 0.05 from one another differ significantly ($P \approx 0.05$ in a Kolmogoroff-test). The data are derived from the 1493 pentamers of table 1.

the length of this unit being strictly conserved. The only deviations within 7.5 kb of satellite DNA analysed are 3 quadrimers. In contrast to this length homogeneity, the repeat unit shows a high degree of sequence variation. In a total number of 1493 pentamers analysed we found 44 variants (tables 1 and 2). The most common of it is 5'TTGCA (V_0), which we term Leitmotiv. We refer to the next most frequent variant 5'TTGAA as V_1 or the main variant and to all remaining variants ($V_2, V_3, \dots, V_{44}$) as minor variants. As indicated by the very different frequencies of variants, mutations (base substitutions) are not randomly distributed throughout the satellite DNA (table 3). Each position within the pentameric repeat unit shows a certain type of predominant base substitution (table 3).

Variants display a disperse distribution pattern (figures 1 and 2, tables 1 and 2) which has the following features: i) Variants do not show local enrichment, as the overall frequencies of the Leitmotiv (60%), the main variant (16%) and the minor variants (24%) hold for all individual satellite DNA fragments analysed. ii) Only the Leitmotiv occurs in tandem arrays, of up to 10 repeats. iii) All variants show a *hyperdisperse* distribution. Assuming that variants arise by chance with probabilities represented by the relative frequencies of their occurrence, the frequency distribution of cluster sizes for a given variant is $\Sigma q^2 p^n = pq$, where $p = p(V_x)$ is the relative frequency of variant V_x , $q = 1-p$ and $n = 1, 2, 3, \dots$ are the cluster sizes. Cluster size 1 means solitary positioning ($V_k V_x V_l$), and cluster sizes $n > 1$ mean n tandem copies of V_x flanked by variants different from V_x ($V_n V_x \dots V_x V_w$). Since $\Sigma n q^2 p^n = p$, the mean cluster size M is $M = q^{-1}$. For example, for variant V_0 (Leitmotiv) with a probability $p(V_0) = 0.6$, we expect relative cluster size frequencies of $0.4(1) + 0.24(2) + 0.144(3) + \dots = 1$ and the mean cluster size $M = 2.5$. However, solitary variant copies (cluster size 1) occur significantly more often than expected by chance (verified by a χ^2 -test). In the case of the Leitmotiv, we found $M = 2.1$, there were 201 solitary copies observed versus 143 expected; for the main variant 237 solitary copies observed versus 168 expected. For the main variant tandem pairs (cluster size 2) occur less often than expected (2 versus 27) and clusters > 2 do not occur at all.

To rule out an unintentional bias for a subpopulation of satellite DNA sequences in the above cloning strategy, we screened two further libraries (for construction see legend to table 1) with satellite DNA and sequenced positive clones. All inserts

BamH I/Mbo I
GGATCCCTTGATTTCATCCTTAAAGCAAGGAAAGAATAACCATTTTAAAGTCCTGTCTACAGATGTGGTAATGGCAAAGTGTATCTCATTTTCG
 GGGCGT GGATC TGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGAA TTGCA TTGCA TTGAA TTGCA ATGCA TTGCA TTGAA
 TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCG TTGAA TTGCA TTGAA CTGCA TTGCA TTGCA TTGCA TTTC A TTGCA
 TTGCA TTGAA TTGCA CTGCA TTGCA TTGAA TTGCA TTGAA TTGCA TTGCA TTGCA TTGAA TTGCA TTGCA TTGCA TTGCA
 TTGAA TTGCT TTGAA TTGCA TTGCA TTGCA TCGCA TTGCA TTGCG TTGAA TTGCA GTGAA TTGCA TTGCA TTGGA TTGCA
 TTGCA TTGAA TTTC A TTACA TTGCA TTGAA TTGCA TTGAA TTGCA TTGAA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA
 TTGCA NNNNN TTGCA TTGCA TTGAA TTGCA TTGCA TTGAA TTGCA TTGCA TTTC A TTGCA TTGCA TTGAA NNNNN TTGCA
 TTGCA TTGCA TTGCA TTGAA CTGCA TTGCA TTGCA TTGAA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA
 CTGCA TTGAA TTGCA CTGCA TTTC A TTGCA TTGCA TTGAA TTGCA CTGCA TTGCA TTGAA TTGCA TTGCA TTGCA TTGCA
 TTGCA TTGAA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGAA TTGCA TTGAA CTGCA TTGAA TTGCA CTGCA TTTC A
 TTGCA TTGCA TTGCA TTTC A TTGCA TTGCA TTGCA TTGAA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGCA
 TTGCA TTGCA TTGAA TTGCA TTTC A TTGCA TTGCA TTGAA TTGCA TTGCA TTGCA TTGCG TTGCA TCGCA TTGCA TTGCA TTGAA TTGCA TTGGA TTGCA TTGCA
 TTGCA TTGCA TTGAA TTGCA TTTC A TTGCA TTGCA TTGAA TTGCA TTGCA TTGCG CTGCA TTGAA TTGCA TTTC A TTGCA TTTC A TTGCA
 TTGC. TTGCA TCGCA TTGCA TTGCA TTGCA TTGCA TTGCA TTGGA TTGCGac

$\frac{1}{2}$ Hinc II

Figure 2. DNA sequence of the telomeric clone pTem1. The telomeric repeats of the somatic genome are pentameric. The BamH I/Mbo I site of the insert, which is joined to the BamH I site of the vector, is underlined and labelled. The fusion of the blunt cloning site of the vector to the telomeric sequences occurs before the underlined sequence denoted $\frac{1}{2}$ Hinc II. The 5 bp consensus sequence at the junction of pentameric repeats and the telomere associated sequence is underlined (see also figure 4). N nucleotide not determined.

correspond to the depicted pentamer type; no further repeat unit was found (nucleotide sequences are not shown, but for pentamer distribution and frequencies see tables 1 and 2). Using cloned satellite DNA as a probe for in situ hybridization (not shown) results in the terminal heterochromatic blocks of the germline chromosomes, which are eliminated by chromatin diminution, being uniformly labelled over their entire length. Thus, at least a substantial amount of the germline limited satellite DNA (highly repetitive DNA) is composed of pentameric sequences.

The telomeric repeats of the somatic chromosomes are pentameric

From previous work, we knew that few copies of the germline limited repeated sequences are present in the somatic genome, where they are located at the chromosome ends (shown by Bal 31 digestion) (31). These copies hybridize to cloned pentameric sequences, suggesting that they are also composed of this type of repeats. In view of the fact that known telomeric sequences are all short tandem repeats (32, 33, 34), it was tempting to speculate that the somatic repeats are not only located very near the chromosome ends, but in fact represent telomeric repeats.

To verify this assumption we constructed two somatic endlibraries and screened them with satellite DNA. We found four positive clones, two out of which (pTem1, library A, pTem5, library B, *telomere megaloccephala*) were sequenced. The insert of pTem1 with a total length of 1175 bp consists of 1070 bp of pentameric repeats in addition to 104 bp of a non-repeated telomere associated sequence (figure 2). The insert of pTem5 with a total length of 3.1 kb was sequenced from both ends of the insert. We determined 530 bp of pentameric repeats and 247 bp of a telomere associated sequence (not shown). Southern blot analysis of various restriction digests of pTem5 using cloned satellite DNA as a probe revealed that the telomere associated sequence makes up at least 1.5 kb of the insert including a 1.3 kb Alu I fragment.

In both clones the Hinc II site of the vector is fused directly to the pentameric repeats indicating that these repeats were in fact terminally positioned in the genomic high molecular weight

DNA fragments. As expected the BamH I and EcoR I site, respectively, of the vector is joined to sequences outside the pentameric repeats.

To show that the pentameric repeats are the sequences of the somatic genome which are sensitive to Bal 31, we measured the kinetics of Bal 31 degradation of high molecular weight somatic DNA. In one experiment the time course DNA samples were cut with EcoR I and in aliquots run in two agarose gels. After Southern blotting one filter was hybridized with cloned pentameric sequences (figure 3A) and the other filter with a clone randomly picked from endlibrary B (figure 3B). In a second experiment the Bal 31 trimmed DNA was cut with Alu I and the Southern blot probed with the insert of pTem5 (figure 3C). In both cases signal intensity and especially fragment sizes of the pentameric sequences decrease dramatically (figure 3A, C), whereas the length of fragments hybridizing with either the random probe or the telomere associated sequence of pTem5 (1.3 kb Alu I band) is unaffected and only a slight decrease in signal intensity is observed (figure 3B, C). This indicates that the high sensitivity of the pentameric sequences to Bal 31 is not due to the general degradation of total DNA. Because the pentameric fragments show a pattern of continuous shortening typical for DNA molecules subjected to exonuclease attack, their sensitivity must be the result of terminal, i.e. telomeric positioning. The possibility that terminal positioning is due to initial Bal 31 cleavage at internal repeats, which might be sensitive per se, is ruled out by the fact that the 1.3 kb Alu I fragment as well as the internal flanking sequence B500 (see figure 4 and results below, marked by an arrow in figure 3B) are not preferentially sensitive to Bal 31.

For clone pTem5 the terminal location was further confirmed by restriction mapping. The end of the telomere can be regarded as a universal restriction site. Somatic DNA was cut with Hind III and after Southern transfer hybridized with the 1.3 kb Alu I fragment of pTem5. This revealed a smear of hybridization typical for the size-variable telomeric fragments.

Thus the telomeric repeats of the somatic chromosomes are pentameric. They show sequence variants in equivalent proportion

Figure 3. The pentameric repeats of the somatic genome are sensitive to Bal 31. A and B: Somatic DNA was digested with Bal 31 for 0 min (a), 0.5 min (b), 1 min (c), 1.5 min (d), 2 min (e), 10 min (f), 30 min (g) and subsequently digested with EcoR I. Time course DNA samples were run in aliquots in two agarose gels (0.8%). After transfer onto nitrocellulose, filter (A) was hybridized with pentameric sequences (1.5 kb insert isolated from subclone pÜ12/B1.5, see figure 4.A), filter (B) with a clone randomly picked from the endlibrary and with the flanking sequence of clone pÜ12 (590 bp BamH I insert isolated from subclone pÜ12/B590, see figure 4.A). The band labelled with the latter probe is indicated by an arrow. C: Somatic DNA was digested with Bal 31 for 0 min (a), 5 min (b), 10 min (c), 20 min (d) and subsequently digested with Alu I. The Southern blot was hybridized with the insert of pTem5. The band lighting up with the telomere associated sequence (Alu 1.3) is indicated by an arrow. Fragment sizes are given in kb. The pentameric sequences are highly sensitive to Bal 31 as signal intensity and fragment sizes decrease dramatically in the course of degradation, whereas fragments hybridizing with the non-pentameric probes are nearly unaffected.

Figure 4. Junction fragments. A: Restriction maps of junction fragments. Solid lines represent flanking sequences; pentameric sequences are boxed. Fragment sizes are given in bp. S Sau 3A, B BamH I, H Hinc II, C consensus sequence GGATC. B: DNA sequence of junction fragments. Shown are about 120 bp of the flanking sequences and two pentameric repeats, respectively. (a) pAm3, (b) pAm8, (c) pTem1, (d) pÜ12, (e) pAm4, (f) pAm1. The 5 bp consensus sequence (in bold letters) is labelled IN 5' and OUT 3' to the pentameric repeats. The frequent trinucleotides TCA and TGA are underlined. N nucleotide not determined.

and distribution as found in the germline limited satellite DNA (tables 1 and 2, for comparison see figure 1).

The junction point between pentameric and flanking sequences is marked by a 5 bp consensus sequence

We sequenced cloned fragments that span the junction between pentameric sequences and others, which we refer to as flanking sequences. If flanking sequences play a role in the breakage process of chromatin diminution and/or amplification of pentamers, they should share common features.

Figure 5. Telomerase based amplification model. In analogy to *Tetrahymena* (18) and *Euplotes* (19) we propose that the telomerase internal RNA in *P. univalens* contains the octanucleotide 3'CGUAACGU of which the trinucleotide 3'CGU functions as binding site (b.s.) and the pentamer 3'AACGU as template site (t.s.). The sequence of both corresponds to the Leitmotiv 5'TTGCA. Most variants are the result of systematic errors in template copying. Considering the most frequent types of base substitution (tables 1 and 3) the following mechanisms appear to operate: i) resulting from transient base stacking changes in the template a base of the t.s. is sequentially presented twice giving rise to V₁(TTGAA), V₂(TTTCA), V₃(TTGGA), V₄(TTCCA); ii) accordingly, the hypothesized C normally adjacent to the t.s. is presented instead of the canonical U giving rise to V₄(TTGCG); iii) prior to correct matching the G of the b.s. is transiently presented in position 1, thus producing V₂(CTGCA). Thus, the predominant types of substitution define the telomerase translocation step. This step is indicated by the frame in which the pentamers are presented. Hybridization of variants to the b.s. is imperfect in most cases. This mismatch can affect the copying precision of the next added repeat. Specifically, binding of V₁ to the internal RNA suppresses the particular copying error which leads to the production of V₁, thus generating the hyperdispersed distribution.

Cloned junction fragments were derived from 4 different libraries. Four clones had been fortuitously found when searching for satellite DNA clones (in the inserts of clones pAm1, 3, 4 and 8). These include flanking sequences of 250 bp, 200 bp, 140/380 bp, and 340 bp, respectively (figure 4A). Clone pÜ12 was found in a library of total germline DNA. It spans 1.5 kb of pentameric sequences (see tables 1 and 2) and a flanking sequence of 2.79 kb (figure 4A). Finally, the telomeric clones pTem1 and pTem5 are junction fragments as well (note that they were derived from somatic libraries). Nucleotide sequences of junction fragments are given in figure 4B.

By hybridizing genomic DNA with the flanking sequences of clone pAm1, pAm4, and pÜ12, respectively, we demonstrated that they are single copy sequences (not shown). We compared somatic and germline DNA and found that the sequence pAm1/250 must become eliminated during chromatin diminution, for it gave no signal with somatic DNA. As the other fragments gave identical hybridization pattern in both DNA types they must originate from the retained fraction.

As can be seen in figure 4B all flanking sequences are AT-rich. Striking are A/T homopolymer runs up to 11 nucleotides. These sequences run differently in agarose and polyacrylamid gels, indicating that they exhibit a sequence directed curvature (35). A second notable feature of the flanking sequences is a frequent occurrence of the trinucleotides TCA and TGA. These trinucleotides are also found frequently in the repeat unit in *A. suum*. Whereas maximal overall homology between the individual flanking sequences does not exceed 52%, all junction fragments share the same 5 nucleotides 5'GGATC immediately at the junction of pentameric and flanking sequences (denoted C in figure 4A, in bold letters in figure 4B). It appears that the 5 bp consensus sequence is an important common feature of junction sites.

DISCUSSION

Here we report that the germline limited satellite DNA of *P. univalens* is, as previously suggested (10), of the simple sequence type, with the following features. The repeat unit is

a pentamer, higher order periodicities not being observed. The pentameric repeat shows a great number of sequence variants; which display a *hyperdisperse* (nonrandom) distribution pattern. How might such an unusual satellite DNA have originated?

The generation of satellite DNAs is widely believed to involve intrachromosomal amplification of a particular sequence during evolution via overreplication and unequal crossing over. For overreplication several mechanisms, such as rolling circle, onion-skin and slippage amplification, have been proposed (7, 36, 37, 38). Homogenization is thought to operate through unequal crossing over and/or gene conversion (39, 40). Variation is well documented for a variety of satellite DNAs from different species (37, 38). In many of these cases variant repeats are arranged in tandem and can be isolated as individual satellites under appropriate conditions, for example in *D. virilis* (41) and *D. melanogaster* (30). In the closely related species *A. suum* there is strict clustering of repeat variants of the germline limited satellite DNA (7). Even where a disperse nonrandom distribution is found, variants tend to cluster into dispersed blocks (42). Thus, homogeneity is a key feature of known satellite DNAs and of proposed mechanisms to explain their generation and dynamics. None of these models can easily explain the hyperdisperse distribution of pentameric variants found in *P. univalens*. Instead we suggest a telomerase based model (figure 5). Support for such a mechanism of amplification comes from the observation that the pentameric repeats have telomeric properties in both genomes. This was first suggested by the observation that experimental fragmentation of the large terminal heterochromatic blocks of satellite DNA in the germline does not lead to breaks that undergo fusion (3, and our unpublished observations), and hence breakage-fusion-bridge-cycles. Thus in this respect the broken ends function as telomeres. In the somatic genome copies of the pentameric repeats are located at the chromosome ends, as demonstrated by their sensitivity to Bal 31 (figure 3) and their occurrence in a somatic endlibrary (figure 2). There is increasing evidence that telomeric repeats in eukaryotes are replicated by telomerases (32, 33). Our model implies that the present sequence diversity of the germline limited satellite DNA of *P. univalens*, containing more than 10^6 repeats per genome, is the direct product of telomerase mediated amplification, mostly unchanged by random mutation. Mechanisms most probably involved in the process of sloppy reverse transcription producing the observed pentameric variants in a nonrandom sequence are briefly suggested in the legend to figure 5. Since the satellite DNA shows no apparent phenotypic expression (43) this suggests a high turnover of this DNA through evolution. The latter idea is consistent with the observation that germline chromosomes show variable amounts of germline limited satellite DNA (44).

The telomeric repeats of *P. univalens* differ from the telomeric sequences of most other species in three respects. They are AT-rich and they do neither display consecutive guanines nor a strand bias for that base. Recently Petracek et al. (34) reported the first example of an AT-rich telomeric repeat in *Chlamydomonas*; but this sequence (TTTATAGGG) still fulfils the two remaining criteria suggested as universal features of eukaryotic telomere sequences (32, 33). It has been suggested that the formation of telomere specific structures occurs via non-Watson-Crick base pairing between guanines (45, 46). We are currently investigating whether the *P. univalens* repeats can function as telomeres in yeast as reported for a number of telomeric sequences from species including human (47).

The *P. univalens* repeats are able to form 'normal' foldback structures, because they are not only direct but nearly perfect

interrupted inverted repeats; and different variants can be complementary to each other. We speculate that the formation of transient foldback structures in the growing single stranded DNA may facilitate the translocation step in the telomerase mediated amplification process.

Sequence analysis of junction fragments reveals a 5 bp consensus sequence, flanked on one side by pentameric repeats and on the other by AT-rich sequences (figure 4). In the case of the somatic endlibrary clones pTem1 and pTem5 the AT-rich sequence represents a telomere associated sequence. AT-rich telomere associated sequences terminating in a short consensus sequence just before telomeric repeats begin have also been found in macronuclear DNA clones of *Tetrahymena* (48, 49). Although telomerase activities have to date only been shown to add telomeric DNA onto telomeric sequences (33), our results suggest DNA ends which are destined to become stable telomeres have special features which allow them to be accepted by telomerase. There are reports of the apparent addition of terminal repeats onto ends that do not carry such repeats in *S. pombe*, *Plasmodium* and most recently in human (50). Indeed, in *Tetrahymena* there is strong evidence that *all* telomeric repeats at macronuclear DNA ends are newly formed. We suggest that after diminutive fragmentation in *P. univalens*, transient somatic chromosome ends are marked by the 5 bp consensus sequence, which serves as a primer for telomerase. We are currently investigating whether the telomere associated sequence of pTem5 is accompanied by pentameric repeats in the germline. If *de novo* addition of telomeric repeats in the somatic genome should be required this would be a further hint toward telomerase mediated amplification in *P. univalens*.

ACKNOWLEDGMENTS

We wish to thank Drs. Wolfram Hörz, David Kipling, and Harry MacWilliams for helpful discussions and critical reading of the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft (Mo 158/17-6,5).

REFERENCES

1. Boveri, T. (1887) *Anat. Anz.* **2**, 688–693.
2. Boveri, T. (1910) *Festschr. R. Hertwig (Jena)* **3**, 131–214.
3. Seidl, C., Bauer, M., and Moritz, K.B. (1988) *Roux's Arch. Dev. Biol.* **197**, 307–320.
4. Beermann, S. (1977) *Chromosoma* **60**, 297–344.
5. Moritz, K.B., and Roth, G.E. (1976) *Nature* **259**, 55–57.
6. Roth, G.E. (1979) *Chromosoma* **74**, 355–371.
7. Roth, G.E., and Moritz, K.B. (1981) *Chromosoma* **83**, 169–190.
8. Streeck, R.E., Moritz, K.B., and Beer, K. (1982) *Nucleic Acids Res.* **10**, 3495–3502.
9. Müller, F., Walker, P., Aeby, P., Neuhaus, H., Felder, H., Back, E., and Tobler, H. (1982) *Nucleic Acids Res.* **10**, 7493–7510.
10. Teschke, C. (1985) *Naturwissenschaften* **72**, 207–209.
11. Goday, C., and Pimpinelli, S. (1986) *Chromosoma* **94**, 1–10.
12. Holmquist, G., and Dancis, B.M. (1980) *Genetica Vols.* **52/53**, 151–163.
13. Moritz, K.B. (1984) *Verh. Dtsch. Zool. Ges.* **77**, 197.
14. Greider, C.W., and Blackburn, E.H. (1985) *Cell* **43**, 405–413.
15. Henderson, E.R., and Blackburn, E.H. (1989) *Mol. Cell. Biol.* **9**, 345–348.
16. Zahler, A.M., and Prescott, D.M. (1988) *Nucleic Acids Res.* **16**, 6953–6972.
17. Greider, C.W., and Blackburn, E.H. (1987) *Cell* **51**, 887–898.
18. Greider, C.W., and Blackburn, E.H. (1989) *Nature* **337**, 331–337.
19. Shippen-Lentz, D., and Blackburn, E.H. (1990) *Science* **247**, 546–552.
20. Shippen-Lentz, D., and Blackburn, E.H. (1989) *Mol. Cell. Biol.* **9**, 2761–2764.
21. Morin, G.B. (1989) *Cell* **59**, 521–529.
22. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor University Press, Cold Spring Harbor.

23. Tartof, K.D., and Hobbs, C.A. (1987) *Focus* 9, 12.
24. Birnboim, C., and Doly, J. (1979) *Nucleic Acids Res.* 7, 1513–1523.
25. Hattori, M., and Sakaki, Y. (1985) *Analyt. Biochem.* 152, 232–238.
26. Feinberg, A.P., and Vogelstein, B. (1983) *Analyt. Biochem.* 132, 6–13.
27. Southern, E.M. (1975) *J. Mol. Biol.* 98, 503–517.
28. Taub, F., and Thompson, E.B. (1982) *Analyt. Biochem.* 126, 222–230.
29. Richards, E.J., and Ausubel, F.M. (1988) *Cell* 53, 127–136.
30. Lohe, A.R., and Brutlag, D.L. (1986) *Proc. Natl. Acad. Sci. USA* 83, 696–700.
31. Teschke, C., Jansen, P., and Moritz, K.B. (1986) *Verh. Dtsch. Zool. Ges.* 79, 332.
32. Zakian, V.A. (1989) *Annu. Rev. Genet.* 23, 579–604.
33. Blackburn, E.H. (1990) *Nature* 343, 122.
34. Petracek, M.E., Lefebvre, P.A., Silflow, C.D., and Berman, J. (1990) *Proc. Natl. Acad. Sci. USA* 87, 8222–8226.
35. Trifonov, E.N., and Ulanovsky, L.E. (1987) In Wells R.D. and Harvey S.C. (ed.), *Unusual DNA structures*. Springer-Verlag, Berlin, pp. 173–187.
36. Beridze, T. (1986) *Satellite DNA*. Springer-Verlag, Berlin.
37. Singer, M.F. (1982) *Intern. Rev. Cytol.* 76, 67–112.
38. John B., and Miklos, G.L.G. (1979) *Intern. Rev. Cytol.* 58, 1–114.
39. Dover, G. (1982) *Nature* 299, 111–117.
40. Jarman, A.P., and Wells, R.A. (1989) *Trends Genet.* 5, 367–371.
41. Gall, G.J., and Atherton, D.D. (1974) *J. Mol. Biol.* 85, 633–664.
42. Jeffreys, A.J., Neumann, R., and Wilson, V. (1990) *Cell* 60, 473–485.
43. Moritz, K.B. (1970) *Verh. Dtsch. Zool. Ges.* 64, 36–42.
44. Moritz, K.B. (1970) In Wied G.L. and Bahr G.F. (ed.), *Introduction to Quantitative Cytochemistry-II*. Academic Press, New York, pp. 57–75.
45. Henderson, E.R., Hardin, C.C., Walk, S.K., Tinoco, I., Jr., and Blackburn, E.H. (1987) *Cell* 51, 899–908.
46. Williamson, J.R., Raghuraman, M.K., and Cech, T.R. (1989) *Cell* 59, 871–880.
47. Cross, S.H., Allshire, R.C., McKay, S.J., McGill, N.I., and Cooke, H.J. (1989) *Nature* 338, 771–774.
48. Yokoyama, R., and Yao, M.-C. (1986) *Nucleic Acids Res.* 14, 2109–2122.
49. Yao, M.-C., Zheng, K., and Yao, C.-H. (1987) *Cell* 48, 779–788.
50. Wilkie, A.O.M., Lamb, J., Harris, P.C., Finney, R.D., and Higgs, D.R. (1990) *Nature* 346, 868–871.