J Mol Evol (1992) 35:114–122

Journal of Molecular Evolution

Telomere-Associated Repeats in *Chironomus* Form Discrete Subfamilies Generated by Gene Conversion

Marita Cohn and Jan-Erik Edström

Department of Molecular Genetics, University of Lund, Sölvegatan 29, S-22362 Lund, Sweden

Summary. In dipteran insects the most distal telomere-associated DNA known to exist consists of long, complex tandem repeats. We have classified the 340-bp tandemly arranged repeats in *Chironomus pallidivittatus*. The repeats are distributed in a small number of subfamilies. One type of the repeat has the character of a master unit from which other main units can be derived usually by simple changes. The derived subfamilies contain segments that are degenerate versions of the corresponding segment in the master sequence. Such segments can also occur together in one and the same repeat unit in different combinations. There is a complete absence of subfamily-specific base variants in regions lying outside of the degenerate segments. Homogenization takes place between DNA sequences that are often smaller than a whole repeat unit. The mosaic structure of the repeat arrays suggests that gene conversion is an important force in the generation and maintenance of this family of repeats.

Key words: Telomeres — Tandem repeats — Subfamilies — Sequence homogenization — Gene conversion — Polymorphism

Introduction

Telomeres from dipteran insects contain long tandem repeats first described in *Drosophila* by Rubin (1978) who suggested that this DNA might be truly terminal, because it occurs in connectives between the ends of the polytene chromosomes. Simple short repeats added by an enzymatic process are known from many eukaryotic telomeres (Blackburn 1991) but have so far not been reported for *Diptera*. Short terminal repeats are not present in mitochondrial DNA molecules from *Tetrahymena*, the ends of which contain complex repeats (Morin and Cech 1988). In the dipteran *Chironomus* the telomere-associated (TA) DNA contains tandem repeats that are 176 bp long (Carmona et al. 1985) or of dimer size, i.e., 340 bp (Saiga and Edström 1985; Cohn and Edström 1991) and 350 bp long (Nielsen et al. 1990). In *Drosophila* the most studied telomere-associated repeat units are the so-called HeT sequences that are much longer and with an irregular tandem arrangement. They are, furthermore, also present in the pericentric heterochromatin (Rubin 1978; Young et al. 1983; Traverse and Pardue 1989). Recently, however, 180-bp TA repeats have also been found in *Drosophila* (Bachmann et al. 1990). In *Chironomus* seven of the eight pairs of telomeric ends, but not the eighth telocentric end, contain TA repeats. The 340–350-bp repeat unit in *Chironomus* is built up of two pairs of subrepeats with high mutual identities and four intervening linker regions, nonrepetitive within the unit. This structure can be understood as a result of selection with varying intensity along the unit, after duplication of a half-unit. This is because mutational differences between units from two species are predominantly located in linker regions (Nielsen et al. 1990).

We have previously characterized two subfamilies of TA repeats from *Chironomus pallidivittatus* (Cohn and Edström 1991). One of the subfamilies has a degenerate stretch with lowered intersubrepeat homology, as if it was derived from the other one by a series of degenerative changes. The larger parts of the sequences of the two subfamilies have, however, homogenized with close to 100% efficiency.

Offprint requests to: M. Cohn

We wanted to learn also whether the remainder of the TA repeat family consists of similar derivatives of the master family. We have, therefore, made a nearly complete characterization of the TA repeat family with regard to subfamily structure. It contains relatively few sequence elements, the prototype (master) sequence with a well-defined subrepeat structure, and a small number of degenerate segments derived from this master sequence. The arrangement of segments allows us to conclude that gene conversion is responsible for homogenizations within and between subfamilies.

Materials and Methods

Cloning, Sequencing, and Screening. Isolated DNA fragments were cloned in pUC18 and sequenced with the dideoxy chain-termination method and ^{35}S-labeled precursor (Chen and Seeburg 1985). Screening of the libraries and Southern hybridizations were made with the previously described clone pCp306, representing a whole repeat unit of telomere-associated DNA (Saiga and Edström 1985). Colony hybridizations and Southern blot analyses were carried out by standard protocols on nylon membranes (Biodyne) (Maniatis et al. 1982). Hybridization conditions were 6 × standard saline citrate (SSC: 0.15 M NaCl, 0.015 M sodium citrate), 0.5% SDS, 5× Denhardt's solution, and 100 μg salmon sperm DNA/ml at 65°C. Final washes for colony hybridization were in 1 × SSC, 0.1% SDS at 65°C and for the Southern analysis 0.1× SSC, 0.5% SDS at 65°C. The oligonucleotides used for screening the libraries are written 5′–3′. The ds2-oligo: ATGGGTGATTCTAGATACTAA. The ds3-oligo: AGCTACGTAAGTAGGCTATCCATC. Screenings with ^{32}P-phosphorylated oligonucleotides were made as above, but at 55°C. Final washes were in 2× SSC, 0.1% SDS at room temperature. Sequences were analyzed with the UWGCG sequence analysis software package (Deveraux et al. 1984).

Results

Structure of the 340-bp TA Repeat

The aim of this study was to characterize as completely as possible the TA DNA with regard to different subfamily variants. In previous experiments we microdissected telomeres for microcloning. The restriction enzyme used was EcoRI. The majority of such clones could be classified into either M1 or D1 340-bp repeats (Cohn and Edström 1991). D1 repeats differ from M1 repeats in having a 30-bp degenerated segment, ds1, mainly localized in the subrepeat IIb (Fig. 1). Because subrepeats show about 90% mutual identity within the repeat unit and because this level is strongly decreased in the affected segment in D1 but not in M1, the D1 unit is considered a derivative of M1. Microlibraries representing several telomeres were screened for the presence of the two subtypes, with oligonucleotides as probes. With a few exceptions, where clones hybridized to both probes, there was a clear hybridization to one or the other of the two oligonucleotides. Although these results did not exclude that there might also be other changes in some of the clones, sequencing of seven clones hybridizing with the M1 oligonucleotide and four clones hybridizing with the ds1-oligo showed them all to be bonafide M1 and D1 units.

Screen for New Subfamilies of TA Repeats

The strategy for screening was to cut genomic DNA with one restriction enzyme and then to eliminate the already characterized repeats with a second enzyme, previously used. The remaining 340-bp band was then isolated and subcloned in pUC18. TA DNA-containing colonies were identified with a clone representing the whole M1 repeat unit, pCp306. EcoRI, HaeIII, and HindII each results in the excision of full-length repeats. Thus, double digests will only result in the release of 340-bp fragments, if one of the sites is mutated. Three different libraries were made in this way, referred to as the HindII, HaeIII, and SnaBI library. D2 and D3 were isolated in this way, and both had a HindII (and a HaeIII) site. D3 has a new restriction site, SnaBI. A double digest with SnaBI and HaeIII allowed the isolation of the variants D4 and D5, which lack the HaeIII site.

HindII Library

Genomic DNA was cleaved with HindII, which cuts most of the TA repeat family (Fig. 2). The HindII site is situated in the middle of subrepeat IIa in the M1 repeat. A second digestion was made with EcoRI to eliminate the M1 and D1 subfamilies, and the remaining 340-bp band was isolated for cloning in pUC18. In total, 467 transformed colonies were collected. Two of the 64 clones that hybridized to pCp306 were randomly selected, sequenced, and compared to the M1 sequence. They differed in a short region including the EcoRI site in L4. This type of unit was designated D2 and the differing segment ds2 (Figs. 1 and 3).

An oligonucleotide was made representing the degenerate segment in D2 and short flanking parts, the ds2-oligo. Twenty-six colonies hybridized to this probe. Two of the 38 remaining TA-positive clones, not hybridizing with the ds2-oligo, were selected for sequencing and compared with the M1 sequence. They were mutually identical and differed from M1 in the whole L4 region. Here a sequence characteristic of L2 was found. In this sequence, furthermore, a short initial segment was present that was a degenerate version of the corresponding sequence in the master version of L2. Because L2 is 1 bp shorter than L4, the whole unit is only 339 bp long. It was

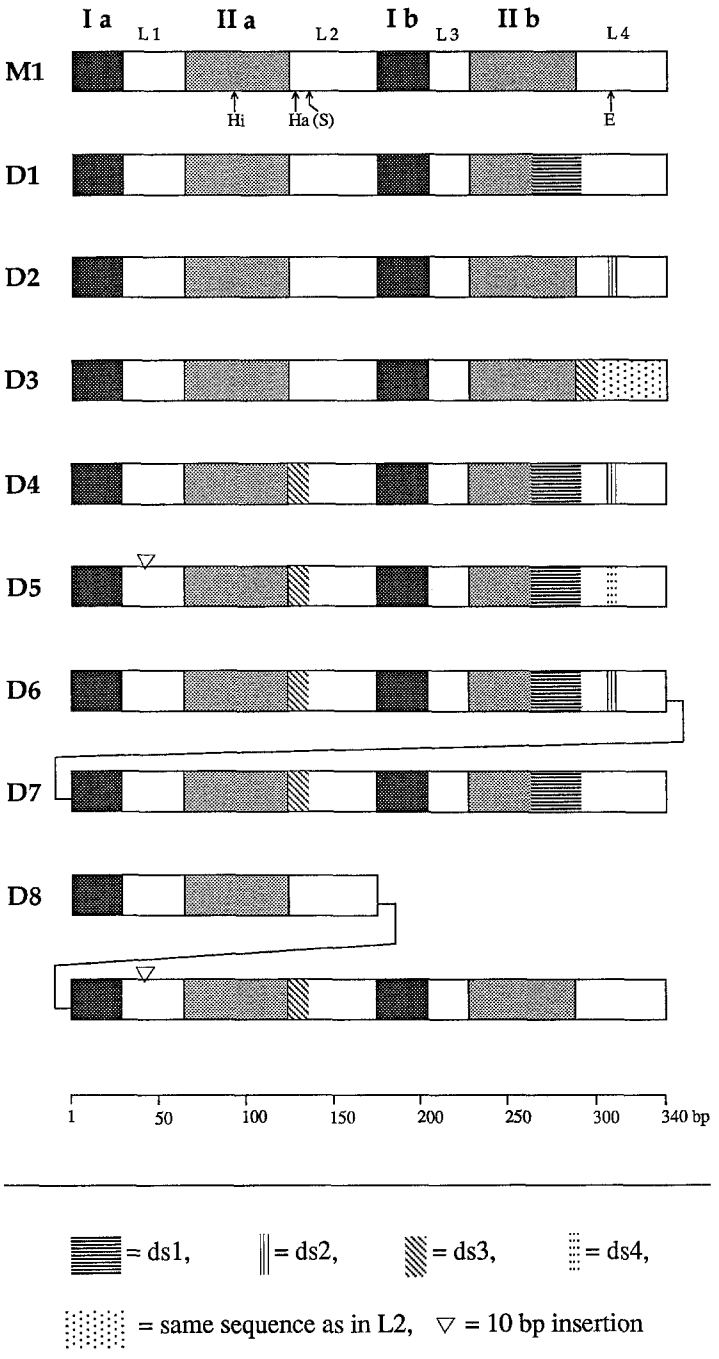

Fig. 1. Schematic representation of the different types of TA repeats, showing the distribution of the derived segments. L1–L4 indicate linker regions. Ia and Ib and IIa and IIb are the two pairs of subrepeats. The connection between D6 and D7 indicates that these two repeats are tandemly arranged in the EcoRI-restricted clone. The sequence downstrem of the cloning site has been transferred from its original place to create repeat units starting with subrepeat Ia. The localization of the four cloning sites has been indicated in M1. Hi = HindII, HA = HaeIII, (S) = SnaBI (only present in ds3), and E = EcoRI.

designated D3 and the short differing segment ds3 (Figs. 1 and 3).

An oligonucleotide, the ds3-oligo, covering the last part of subrepeat IIb and the ds3 derivate, was then used to screen the remaining clones. Thirty-one ds3-containing clones were identified. With a few exceptions (seven) all the HindII-cleavable, EcoRI-resistant units belonged to the D2 or D3 subfamilies. Three of the exceptions, which all showed weak hybridization with pCp306, were sequenced but were not TA repeats.

Like ds1, the ds3 segment is likely to be a degenerate version of the corresponding part in M1. The motif is 10 bp long with five differences from the M1 segment. The identities of the two M1 original homologues in L2 and L4 is 7 out of 10 positions. The ds3 motif shows, however, only agreement in 2 of 10 positions with the original homologue in L4 indicating that it has accumulated mutations more rapidly than the M1 segment.

HaeIII Library

In order to select for additional diverged repeats, the procedure was repeated with HaeIII as the first enzyme. HaeIII, which has a recognition site in L2, also cleaves most of the TA DNA to 340-bp units (Fig. 2). Genomic DNA cleaved with HaeIII was

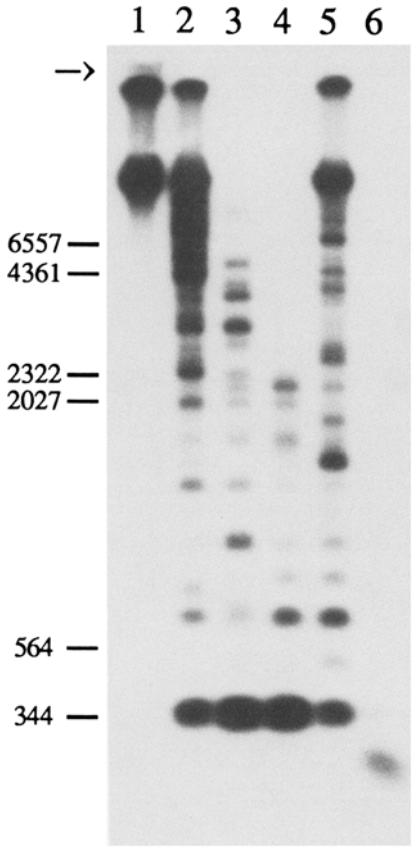

Fig. 2. Southern blot analysis of genomic DNA from *C. pallidivittatus*. Each lane contains 1 μg of DNA cleaved with the restriction enzymes (1) HindIII, which does not cleave the TA repeats; (2) EcoRI; (3) HindII; (4) HaeIII; (5) SnaBI; and (6) DdeI, which cleaves TA repeats to half-units. The separations were made in 1% agarose gel.

subsequently treated with HindII and EcoRI. The cleavage with HindII only reduces the HaeIII fragment with about 35 bp. When DNA fragments in the 340-bp region were eluted, some DNA cleaved with both HindII and HaeIII could therefore have been included. Six hundred twenty-four transformants were collected. In order to eliminate already known repeats, the 33 clones hybridizing with pCp306 were further screened with the two oligonucleotides ds2 and ds3. Almost all the TA DNA-containing clones also hybridized to one or the other of the two oligonucleotides (17 to the ds2-oligo and 12 to the ds3-oligo). Four clones remained negative, but these clones hybridized only weakly with pCp306. Sequencing analysis of one such clone showed that it did not contain true TA DNA. Hence, this library did not give any additional derived segments.

SnaBI Library

The ds3 derivative generates a SnaBI site, but it also eliminates a preexisting unique HaeIII site in L2. In the D3 subfamily the nonsubstituted L2 is present in addition to the L2 segment containing the ds3 motif in the position of L4. The D3 units are, therefore, cut with both enzymes. We were interested in knowing whether the ds3 motif might occur in the L2 sequence, also when the L2 sequence occupies its original position. Such units would be resistant to HaeIII but be cut by SnaBI. We obtained support for this view when we found that the fraction of TA DNA that still remains at larger than monomer size after the combined restriction with EcoRI, HindII, and HaeIII is cleavable with SnaBI (Fig. 4). In order to eliminate the D3 subfamily we therefore cleaved genomic DNA with both SnaBI and HaeIII and isolated and cloned the 340-bp band. Six hundred twenty-four clones were initially collected, and 91 of them were positive to pCp306. These clones were screened with the ds3-oligo. Three hybridizing clones were obtained and subsequently sequenced. They all had ds3 in L2 in its normal position but they also contained more than one degenerate region. Two of the sequences were mutually similar and are designated D4 in Figs. 1 and 3. This unit contains, in addition to ds3, the two other derivatives ds1 and ds2. The third sequence, named D5, contains both ds1 and ds3, but lacks ds2. The EcoRI site is here lost due to a new kind of derived segment not observed before, ds4. D5 also contains a 10-bp insertion in L1. In conclusion this screen did not recover repeat units containing the ds3 region in an M1 background, but it resulted in the isolation of a couple of complex units, D4 and D5. The latter is atypical in several respects. The derived segment ds4 has been observed only once, and the repeat contains no less than six base substitutions, three insertions, and two deletions outside of the three degenerate segments, ds1, ds3, and ds4.

A Double, Complex Unit

Having obtained the surprising result that repeat units may contain more than one degenerate segment, we turned to the already investigated EcoRI-restricted and microcloned library of telomere 3L (Cohn and Edström 1991). This library contained a few not yet characterized clones. A clone containing a 680-bp insert was sequenced. It contained two repeats in tandem, where the internal EcoRI site had been lost. This fragment showed a very complex pattern with respect to the derived segments (Figs. 1 and 3, D6–D7). Three derived segments, ds1, ds2, and ds3, were present. The ds3 segment was found in the L2 region, here occupying its original place as in D4 and D5.

The 1.5-mer Unit

A ds3 region in L2, as in D4–D7, has been observed before in a 525-bp insert cleaved with EcoRI (Saiga

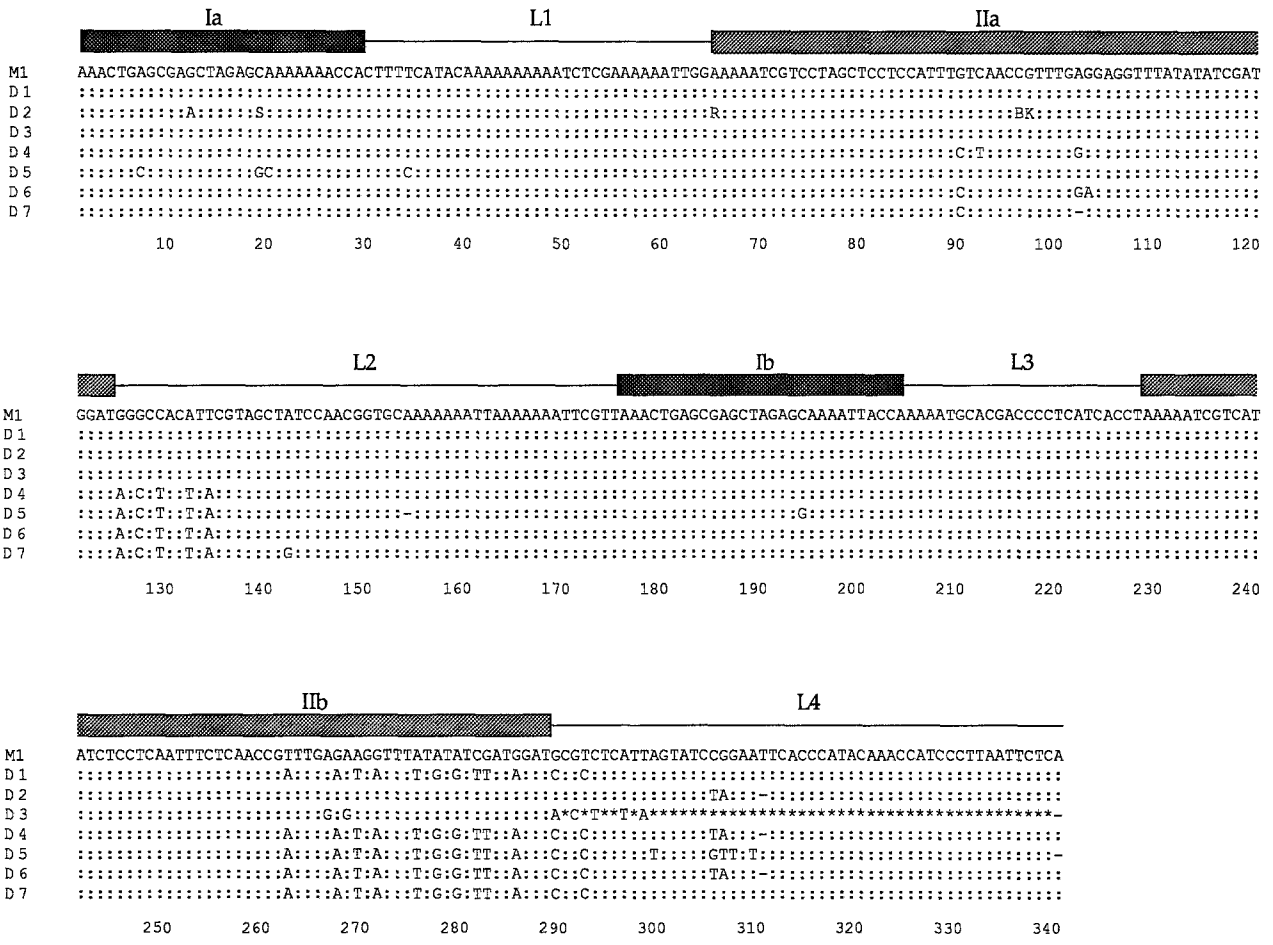

Fig. 3. Sequences of the different types of TA repeats found in *C. pallidivittatus*. M1 is the master ssequence, representing the consensus sequence produced from seven different clones (Cohn and Edström 1991). A colon (:) refers to base identity with the M1 sequence and hyphen (-) a gap. An asterisk (*) indicates base identity with the L2 sequence in M1. Uncertainties in base determination are labeled R = A or G; S = C or G; K = G or T; and B = C, G, or T. D1 represents a consensus sequence of four clones (Cohn and Edström 1991). M1 and D1 are both derived from EcoRI microclones. D2 represents two clones from the HindII library. One of the sequences is shown here. The other sequence differs in two places; G in position 12 and an insertion of A between bases 39 and 40. D3 represents two identical clones from the HindII library. D4 represents two clones from the SnaBI library. The clone not shown has a G in positions 19 and 142. D5 is a single clone from the SnaBI library, with insertions between the following positions: 43 and 44, TTTCCCAAAA; 138 and 139, G; 149 and 150, CA. D6 and D7 together represent one single 678-bp insert from a microclone library.

and Edström 1985). This insert corresponds to one and a half repeat units in length and is listed as D8 in Fig. 1. Interposed between the normal A and B halves of the M1 unit, there is a modified A unit in which the L2 region contains a ds3 derivative. There is, furthermore, the same 10-bp insertion in L1 that is found in D5. This insertion also is present in TA repeats from a sibling species, *C. tentans* (Nielsen et al. 1990). The D8 unit has only been found once, but no systematic search has been done. It is probably unusual, as a complete digestion with EcoRI gives a hardly visible band at the 525-bp position (Fig. 2).

Nonrandom Base Substitutions and an Insertion

Apart from the four segments in which the M1 sequence was changed there were some other nonrandom variants in our collection of repeats. Thus, C is changed to G in position 19 in D5 and one of the D4 clones. The 10-bp insertion between positions 43 and 44 in D5 and D8 is also present in *C. tentans* (Nielsen et al. 1990) The change of G to C in position 90 is present in D4, D6, D7, and D8. This substitution eliminates the HindII site and explains why not all TA DNA is digested to monomers by HindII. The change A to G in position 102 occurs in D4, D6, and D8. A is changed to G in position 142 in D7 and one of the D4 clones. The T in position 299 in D5 has also been found in one of the sequences constituting the D1 consensus (Cohn and Edström 1991). Finally, in the two D3 clones, A is changed to G in positions 266 and 268, which makes this region is subrepeat IIb identical with the corresponding region in IIa for one of the clones containing ds3 in the original position (D4). This sug-

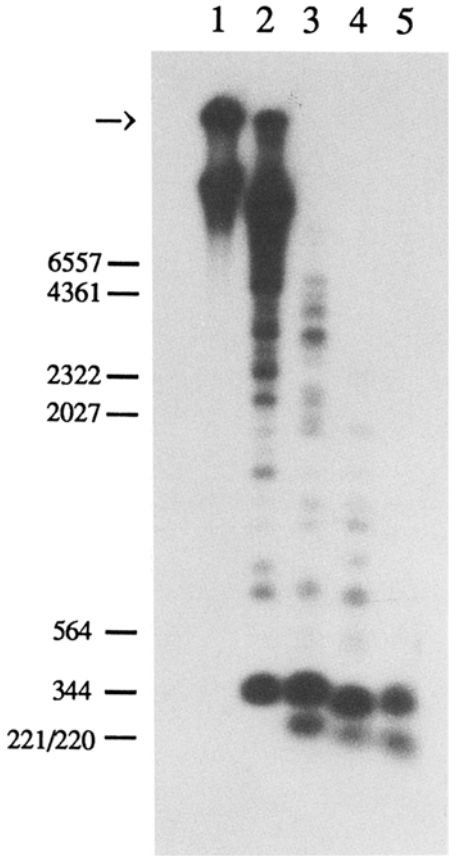

Fig. 4. Genomic blot showing the amount of TA DNA, which is made low molecular weight with the restriction enzymes used in the cloning procedures. The final step shows that no high molecular weight DNA is left after digestion with all of the enzymes used. In each lane 1 μg of genomic DNA is separated on a 1% agarose gel. Lane 1 shows undigested DNA, then follows digestions with (2) EcoRI; (3) EcoRI and HindII; (4) EcoRI, HindII, and HaeIII; and (5) EcoRI, HindII, HaeIII, and SnaBI.

gests that D3 is formed by a recombination event in which not only L2 with an included ds3 but also part of the preceding IIa subrepeat has participated.

Consequently, the family of TA repeats is polymorphic not only with regard to the presence of degenerate motifs, but also with regard to an insertion and some base substitutions.

Completeness of Screen

Figure 2 shows that EcoRI and SnaBI both liberate similar amounts as monomers (about 30% of the total TA DNA, data not shown) and that they both produce more high molecular weight fragments than HindII and HaeIII. We estimate that the release of monomers is between 80 and 90% for HindII and HaeIII. The blot also shows that the distribution of sites is far from random.

The course of digestion with these four enzymes can be followed step by step in Fig. 4. It shows the absence of material larger than monomers after the last step and allows us to conclude that our screen must have included all quantitatively important subfamilies.

Our experience from the application of several other restriction enzymes confirms that there is very little sequence variation apart from the one we have seen. Several enzymes with four-base recognition sites, that are not present in any of the repeats isolated, do not give measurable degradation of TA DNA (data not shown). Application of enzymes for sites present in regions where we find little if any variation, give no measurable amounts of material larger than the monomer. One example is ClaI with a constant site in IIa and IIb, except for the ds1-containing subrepeats (Saiga and Edström 1985; Nielsen et al. 1990). Also DdeI has sites in both Ia and Ib that are highly constant (Fig. 2).

We conclude that M1, D1, D2, and D3 constitute the main important subfamilies present in the TA DNA. In addition to these, there is only a very small number of more complex units. This is also supported by the screenings performed on the HindII and HaeIII library with the ds2- and the ds3-oligo. Only one single clone (not sequenced) from these libraries hybridized to both oligonucleotides. M1 is probably more common than D1 (Cohn and Edström 1991), whereas D2 is about as common as D3 (according to the screenings), but all four subfamilies are quantitatively important.

Distribution of Derived Segments

The small and largely predictable variability in the TA repeat family provides a highly convenient means of studying the distribution of different derived regions in the arrays of TA repeats. Thus, Sau3AI is specific for ds1, HinfI for ds2, and SnaBI for ds3. M1 shares the EcoRI site with D1 and the SfaNI site with D2 (among the main subfamilies). Figure 5 shows that ds1 is to a large extent interspersed, as the monomer band is weak after application of Sau3AI. Most of these sites are separated by more than one repeat unit, among which there are also a few of the 1.5-mer size, probably D8 units. After digestion with HinfI, diagnostic of ds2, the distribution is very different with a larger fraction of tandemly arranged units. SnaBI (ds3) gives a pattern containing all of the half-mer steps, which is to be expected as ds3 can occur in both halves of the 340-bp repeat. The pattern shows that many of these units are interspersed with others. SfaNI and EcoRI, finally, have strong monomer bands and both have characteristic, strong 2.5-mer, 4-mer, and 6-mer bands. The traits characteristic for both should reflect the distribution of M1 units, which are consequently also partially interspersed. The strong monomer bands indicate, however, that the M and

D2 subfamilies to a large extent are tandemly arranged. All the enzymes show higher level bands that are very different in intensities. This suggests periodicities in the distribution of the different units, characteristic for each type.

Discussion

We have assessed much of the variability in a large family of repeat units, the telomere-associated repeats in *Chironomus*. The units belong to different well-defined subfamilies, within which the sequence agreement is very high, as a rule well below the 1% base substitutional level.

Among the different subfamilies one has the character of a master unit from which the others can be derived, usually by a single differing segment. Here the derived sequence has undergone a series of mutational changes. Alternatively, there is a substitution with a sequence from another region of the repeat unit containing a mutated patch. In regions lying between degenerate segments, there is a close to 100% sequence agreement between master and derived units.

By all evidence, we have obtained the main sequence variants; M1, D1, D2, and D3. Each derived unit has a characteristic degenerate motif: ds1, ds2, and ds3, respectively. D3 is different from the other derived subfamilies in having a structure that suggests that it is a result of unequal crossing-over within the repeat unit. L2 with its derived segment ds3 here occupies the position of L4. In addition we obtained some rare units that contain two or three motifs. They are informative because they show that the recombining unit during sequence homogenization must be smaller than the repeat unit. Finally, there is a class of derived units in which a large duplication is combined with a degenerate patch (D8).

Although it may be convenient to classify the TA DNA in terms of 340-bp repeat units, this does not quite truly reflect the structure of this family. It is probably more accurate to consider it as a basic matrix of tandemly arranged M1 units into which different degenerate motifs are inserted relatively independently of each other. As a consequence, a ds segment can be lacking in a repeat (M1) or occur alone (D1, D2, and D3). More rarely, however, two (D7) or three motifs may occur in a single unit (D4, D5, D6). Such a model is also in agreement with the blotting experiments with motif-specific enzymes. They show that only a fraction of repeats, varying in size for different units, are lying in tandem.

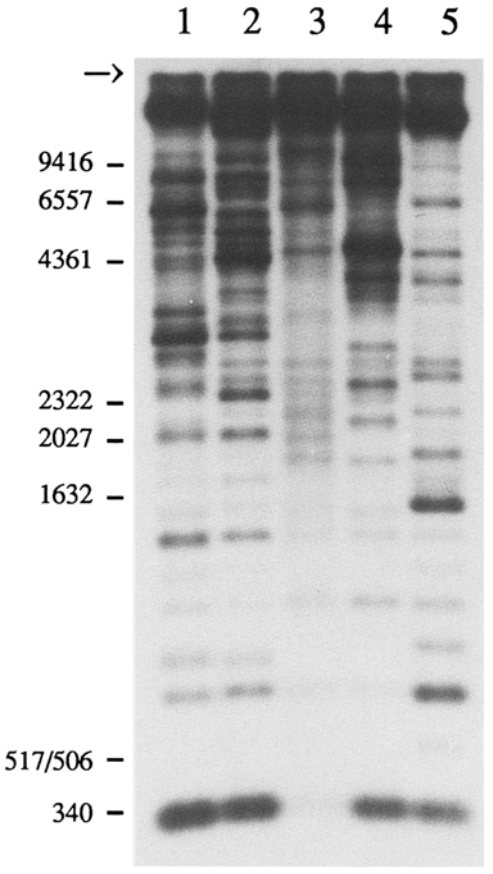

Fig. 5. Arrangement of the derived segments in TA DNA as shown by restriction of genomic DNA with enzymes specific for each segment. Each lane contains 1 μg DNA cleaved with (1) SfaNI, (2) EcoRI, (3) Sau3AI (ds1), (4) HinfI (ds2), (5) SnaBI (ds3), and separated on a 1% agarose gel. The recognition site of SfaNI is situated in the border between subrepeat IIb and L4 like Sau3AI. EcoRI and HinfI are in L4 and SnaBI in L2.

Mechanism of Homogenization

The TA repeat family, as a whole, is arranged in long regular tandem repeats (Saiga and Edström 1985). Such a structure is likely to result from unequal crossing-over. This process is, however, not sufficient to explain the ongoing homogenization between the repeat units. Our results show that the participating units often must be smaller than the repeat unit. This suggests that gene conversion is necessary for maintaining homogeneity (Dover 1987).

The ds regions are created in a continuous process (although not necessarily at a uniform rate) representing accumulations of mutations in ancestral M1 segments. Although unequal crossing-over could certainly have a homogenizing effect in the sense that it might increase or decrease the numbers of members of a subfamily or ds region, it is nevertheless not likely that it has an important role in the formation and maintenance of the ds regions. If this was the case, an expected effect would be the oblit-

eration of the sharp mosaic structure of the derived repeat arrays. This is because there is no reason why unequal crossing-over, acting within a subfamily, would spread new mutations only within ds regions.

Mosaic traits assumed to result from gene conversion events have been observed before in repeat families (Weiss et al. 1983; Strachan et al. 1984, 1985; Hughes and Nei 1989). The TA repeat family, however, appears quite exceptional in this respect, underlining the prominent role of gene conversion in the evolution of this family. Patterns of variation in minisatellite families (Jeffreys et al. 1990) also suggest a major role for gene conversion for their evolution. Finally, it has been found that ribosomal gene arrays evolve by gene conversions in triploid parthenogenetic species with parental chromosome sets of different origin (Hillis et al. 1991).

Types of Discontinuous Variation

How can the pronounced discontinuous variation in the TA repeat family be explained? One factor that can give rise to subfamilies is chromosomal localization, as shown for alphoid centromeric repeats (Willard and Waye 1987). In the TA repeat family, however, members of one subfamily are present in different telomeres, and a given telomere may contain members of more than one subfamily (Cohn and Edström 1991).

Other examples of discontinuous variation are the gene–pseudogene families, reminiscent of the present one in the sense that there may be regions of high identity interrupted by segments of low or lacking identity (Miyata and Yasunaga 1981). If we assume that the M1 subfamily is the gene equivalent, the D subfamilies would correspond to pseudogenes. TA repeats are transcribed, but telomeres are not transcribed to an equal extent (Carmona et al. 1985; Botella et al. 1991). The TA subrepeat structure, furthermore, is likely to be a result of selective forces, subrepeats being evolutionarily conserved regions (Nielsen et al. 1990). At least the two segments ds1 and ds3 are obvious degenerate versions of the corresponding segments in M1.

The gene–pseudogene paradigm has the added advantage that it could explain the remarkably low base substitutional level within each subfamily. This could be due to selection against nucleotide changes combined with frequent gene conversions. D subfamilies would, according to this view, be corrected after the M family except in the ds segments where there is a block against such a correction, explaining their accumulation of base substitutions. Each ds segment, nevertheless, is homogenized within its own class with the same high efficiency; so far there has been a 100% agreement within ds classes in our analyses.

Correlation between End-to-End Contacts and TA Sequences

Rubin (1978) obtained evidence for a subfamily structure among telomere-associated repeats in *Drosophila* and speculated that they might serve a role in telomere recognition. It has been observed that the frequencies differ for different intertelomeric contacts in *Drosophila* (Hinton 1945; Warters and Griffen 1950). Alternatively, such differences could be explained by the presence of a specific repeat sequence with a differential intertelomeric distribution (Renkawitz-Pohl and Bialojan 1984).

The TA repeats in *Chironomus* are present in seven of the eight pairs of telomeres (Carmona et al. 1985; Saiga and Edström 1985). These can undergo mutual end-to-end contacts, but the eighth, telocentric pair is excluded from such contacts (Bauer 1936). The members of this pair may, however, show mutual association, which can be demonstrated when the somatic pairing of the two homologues is inhibited (Beermann 1955). We have previously found that although the identities between members of a certain subfamily are the same within and between telomeres, chromosome ends may contain different proportions between members of a certain subfamily (Cohn and Edström 1991). It is, however, not known whether these end-to-end contacts are functionally meaningful or are just passive consequences of mutual sequence similarities. In order to approach the question regarding the functional role of TA repeats, we need, among others, information about the precise localization in the telomeres of the TA variants and their representation in the telomeric transcript. Such work is in progress.

Acknowledgments. This work was supported by grants from the Swedish Cancer Society and the Philip–Sörensen Foundation.

References

Bachmann L, Raab M, Sperlich D (1990) Evolution of a telomere associated satellite DNA sequence in the genome of *Drosophila tristis* and related species. Genetica 83:9–16

Bauer H (1936) Beiträge zur vergleichenden Morphologie der Speicheldrüsenchromosomen (Untersuchungen an den Riesenchromosomen der Dipteren. II). Zool Jahrb (Physiol) 56: 239–276

Beermann W (1955) Cytologische Analyse eines Camptochironomus-Artbastards. I. Kreuzungsergebnisse und die Evolution des Karyotypus. Chromosoma 7:198–259

Blackburn EH (1991) Structure and function of telomeres. Nature 350:569–573

Botella LM, Morcillo G, Barettino D, Diez JL (1991) Heat-shock induction and cytoplasmic localization of transcripts from telomeric-associated sequences in *Chironomus thummi*. Exp Cell Res 196:206–209

Carmona MJ, Morcillo G, Galler R, Martinez-Salas E, de la Campa AG, Diez JL, Edström J-E (1985) Cloning and molecular characterization of a telomeric sequence from a temperature-induced Balbiani ring. Chromosoma 92:108–115

Chen EY, Seeburg PH (1985) Supercoil sequencing: a fast and

simple method for sequencing plasmid DNA. DNA 4:165–170

Cohn M, Edström J-E (1991) Evolutionary relations between subtypes of telomere-associated repeats in *Chironomus*. J Mol Evol 32:463–468

Deveraux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395

Dover GA (1987) DNA turnover and the molecular clock. J Mol Evol 26:47–58

Hillis DM, Moritz C, Porter CA, Baker RJ (1991) Evidence for biased gene conversion in concerted evolution of ribosomal DNA. Science 251:308–310

Hinton T (1945) A study of chromosome ends in salivary gland nuclei of *Drosophila*. Biol Bull 88:144–165

Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc Natl Acad Sci USA 86:958–962

Jeffreys AJ, Neumann R, Wilson V (1990) Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. Cell 60:473–485

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor University Press, Cold Spring Harbor NY.

Miyata T, Yasunaga T (1981) Rapidly evolving mouse α-globin-related pseudogene and its evolutionary history. Proc Natl Acad Sci USA 78:450–453

Morin GB, Cech TR (1988) Mitochondrial telomeres: surprising diversity of repeated telomeric DNA sequences among six species of *Tetrahymena*. Cell 52:367–374

Nielsen L, Schmidt ER, Edström J-E (1990) Subrepeats result from regional sequence conservation in tandem repeats in *Chironomus* telomeres. J Mol Biol 216:577–584

Renkawitz-Pohl R, Bialojan S (1984) A DNA sequence of *Drosophila melanogaster* with a differential telomeric distribution. Chromosoma 89:206–211

Rubin GM (1978) Isolation of a telomeric DNA sequence from *Drosophila melanogaster*. Cold Spring Harbor Symp Quant Biol 42:1041–1046

Saiga H, Edström J-E (1985) Long tandem arrays of complex repeat units in *Chironomus* telomeres. EMBO J 4:799–804

Strachan T, Sodoyer R, Damotte M, Jordan BR (1984) Complete nucleotide sequence of a functional class I HLA gene, HLA-A3: implications for the evolution of HLA genes. EMBO J 3:887–894

Strachan T, Webb D, Dover GA (1985) Transition stages of molecular drive in multiple-copy DNA families in *Drosophila*. EMBO J 4:1701–1708

Traverse KL, Pardue ML (1989) Studies of He-T DNA sequences in the pericentric regions of *Drosophila* chromosomes. Chromosoma 97:261–271

Warters M, Griffen AB (1950) The telomeres of *Drosophila*. J Hered 41:182–195

Weiss EH, Mellor A, Golden L, Fahrner K, Simpson E, Hurst J, Flavell RA (1983) The structure of a mutant H-2 gene suggests that the generation of polymorphism in H-2 genes may occur by gene conversion-like events. Nature 301:671–674

Willard HF, Waye JS (1987) Hierarchical order in chromosome-specific human alpha satellite DNA. Trends Genet 3: 192–198

Young BS, Pession A, Traverse KL, French C, Pardue ML (1983) Telomere regions in *Drosophila* share complex DNA sequences with pericentric heterochromatin. Cell 34:85–94

Received February 3, 1992/Revised March 17, 1992