Nucleic Acids Research, 1993, Vol. 21, No. 8 1775–1781

A repetitive DNA sequence associated with the centromeres of *Chironomus pallidivittatus*

Carlos Rovira, Wolfgang Beermann[1] and Jan-Erik Edström*
University of Lund, Department of Molecular Genetics, Sölvegatan 29, S-223 62 LUND, Sweden and
[1]Eduard Spranger-Strasse 55, D-7400 Tübingen 1, Germany

Received January 1, 1993; Revised and Accepted March 19, 1993 EMBL accession no. X66064

ABSTRACT

A clone containing centromere-associated DNA from *Chironomus pallidivittatus* was obtained by microdissection-microcloning. It hybridizes to the centromeric end of one chromosome and exclusively to regions in the three remaining, metacentric chromosomes to which centromeres have previously been localized on cytological grounds. In the metacentric positions the hybridization can be assigned to thin bands. The clone contains 155bp tandem repeats and short flanking regions represented in all of the centromeres. Titration experiments show that the four centromeres together contain 200kb of 155bp repeat per genome. In a line of tissue culture cells the amounts are increased by a factor 1.5–2, resulting in proportionately extended arrays of tandem repeats. Each repeat contains two invertrepeats surrounding a region containing only AT base pairs, a feature with some similarity to functionally essential elements in the *Saccharomyces cerevisiae* centromere.

INTRODUCTION

Centromeres have been defined structurally and functionally in *Saccharomyces cerevisiae* (1) and *Schizosaccharomyces pombe* (2–4), in the former case as a relatively simple unit, in the latter case as a complex structure. In higher eukaryotes, however, only certain aspects of centromere structure have been defined at the molecular level. This, among others, is due to the complexity and large size of the chromosomal regions in which centromeres are localized.

Different blocks of highly repetitive DNA sequences representing several percent of the mammalian genome hybridize *in situ* predominantly to the centromeric heterochromatin and consist of long tandem arrays of repeat units (5, 6). In primates a dominant component of centromeres is the tandemly arranged α-satellite, consisting of a 171bp repeat sequence organized in higher-order, chromosome specific repeat units in blocks in the 500–5000 kb range (7, 8). Also other classes of repeated DNA sequences (satellites II+III) have been found associated with human centromeres. No function has been assigned to these other classes, although it has been suggested that they have traits indicative of centromere function (9). In other mammals the dominant centromeric repetitive DNA component may look different. The mouse centromere has a 120bp minor satellite sequence which is unrelated to the α-satellite of humans except for a conserved 17bp portion (10, 11). Several proteins localized in the centromeric region were detected with the help of autoantibodies present in the serum of patients with a rheumatic disease known as the CREST-syndrome (12). One centromere binding protein (CENP-B) has been shown to bind to a subset of α-satellite monomers containing the conserved 17-bp part, termed the CENP-B box, after the binding properties (13). Recently, it has been shown that CENP-B binds to human α-satellite DNA introduced into African Green Monkey cells and that the presence of transfected DNA has effects on chromosome segregation during mitosis, suggesting that α-satellite is responsible for at least some aspects of centromere function at the DNA level (14).

In *Drosophila* the transcriptionally inactive satellite DNA represents up to 25% of the genome (15). Four different satellite species have been found. They exist in large blocks and are localized in the centromeric region of the three autosomes, the proximal heterochromatin of chromosome X and along all of the heterochromatic Y-chromosome. The function of these satellites is unknown.

In the dipteran genus *Chironomus* several families of satellite like, tandem repetitive sequences have been reported (16). Among others a short repetitive DNA element defined by a *Cla*I restriction site (*Cla*I-DNA family) hybridizes *in situ* predominantly to the centromeric heterochromatin in various chironomid species but also to interstitial regions in the chromosome arms. Its function is unlikely to be related to control of cell division.

Whatever its molecular architecture, the eukaryotic centromere is part of regions containing several hundred or even thousand kilobases of DNA. This large amount of DNA has impeded the characterization of the eukaryotic centromere at the DNA level. Chironomids that are members of the subgenus *Camptochironomus* may, however, represent organisms with a much more favourable centromeric size. Thus centromeres have

* To whom correspondence should be addressed

tentatively been assigned to relatively thin bands in the three large metacentric (somatically paired) polytene chromosomes and at the left telomere of the small chromosome four by Beermann (17).

In this report we present a novel repetitive sequence from *Camptochironomus* which hybridizes specifically to the centromeric regions of *C. pallidivittatus*. The hybridization signals in the three large chromosomes are lying in thin and well defined bands which correspond to a titrated amount of about 200kb for the four centromeres together. Some structural features of this putative centromeric sequence are similar to functionally essential centromere elements in *Saccharomyces*. Also a piece of DNA flanking the tandem repeat arrays, present in the four centromere pairs is described. Access to this DNA will be of importance for a more detailed structural characterization of the four centromere pairs. Whereas our data suggest that arrays of centromeric tandem repeats are retained in size during polytenization they also indicate a considerable amplification in a tissue culture line.

MATERIALS AND METHODS

Microdissection, microcloning, screening and sequencing procedures

Microcloning was done essentially as described by Scalenghe et al (18). The centromeric end of chromosome IV (19) was dissected with a de Fonbrune micromanipulator. Two pieces were collected for the cloning experiment. The material was digested with restriction endonuclease *Eco*RI and ligated in λgt-10. From a collection of several hundred recombinant clones, 72 plaques were individually isolated and screened with nick-translated total genomic DNA (20). The 627bp insert from one of two positive clones was subcloned into pUC18 for further analysis and named pCp627. The sequence was determined by the dideoxy chain termination method.

```
1   GAATTCAGAC AAAACAGAAC ATCAGACTTA ACTTCTCATA TATTCCCTTT
51  GAAAATATTA TCTCTCTAAG CGAGCGAATG CTTTACAACC GTTGAAAAAA
101 GAGTGGTTCA TATATATTGG CATTCTGCCA ATGCCAATGG CAAAGCCCGA
151 TAGCTCAGTG GTCTGAGCAC TTGACGGCAA TCGAGAGGTG CTAGGTTCGT
201 TCCCGCTCGG GAAGAGTCAT TGGGTGAATT ACTTTTTTTT CAACTTTTTC
          |------repeat 1------>
251 TTTTAAAGCT TTTATTTTAT TTAATTTTTT TTAAAGGCTT TGGAGCAAAT
    <------>                      <------>
301 CTTTTAATCA GTCATAATTT TTATGTGGTC TCTGTGGCAA TATCGCCCAT
          ^ ^
351 AGAACATTTT TTTATTTGAG CCTTGGAAAG CATTTTTTGA CAATTTTTTC
    ------>|<-- repeat 2 ------------------------>
401 ATTTTTCAAA AAAGCTTTTA TTTTATTTAA TTTTTTTTAA AGGCTTTGGA
    <------>                      <------>
451 GCAAATCTTT TAGTCGCATA ATTTTTATAT GGTCTCTGTG GTAATATCGC
501 CGATAGAACA TTTTTTGATT TGAGCCTTGG AAAGCATTTT TTGACAATTT
          ------>|
551 TTTCATTTTT CAAAAAAGCT TTTATTTTCA AAGGAATATA TGAGAAGTTA
    <------>
601 AGTCTGATGT TCTGTTTGTC TGAATTC
```

Figure 1. Nucleotide sequence of the clone pCp627 containing the two tandemly repeated sequences preceded by 254bp flanking sequence and followed by 54bp that are an inverted version of the first 54bp of the clone. The pairwise invertrepeated structures located at the beginning of each repeat are underlined by bidirectional arrows. Base pair differences between the two tandem repeated sequences are marked by dots. Bases not present in the second repeat are marked by arrows.

Preparation of high molecular weight DNA

Epithelial tissue culture cells were pelleted by centrifugation, washed twice in PBS and resuspended at a concentration of 3×10^8 cells per ml. The cell suspension was then mixed with an equal volume of 1% agarose dissolved in sterile water at 50°C. The mixture was pipetted into plastic molds. After the agarose had settled, DNA preparation and restriction endonuclease digestion was carried out as described by Sambrook et al (21).

The larval high molecular weight DNA was isolated as described by Bingham el al (22) with minor modifications.

Pulsed-field gel electrophoresis (PFGE)

The blocks containing tissue culture cell high molecular weight DNA were inserted into 1% agarose gel. The high molecular weight larval DNA was pipetted directly into the wells. The samples were separated by PFGE with a Gene Navigator unit (Pharmacia-LKB) in 0.5×TBE buffer at 9°C. The power supply was set at 300V for 13 hrs. The pulse time was increased stepwise as follows: 1hr 0.3 sec, 1hr 0.5 sec, 1hr 0.7 sec, 5 hrs 2.0 sec and finally 5hrs 4.0 sec. After separation the gel was stained by immersion in distilled water containing ethidium bromide (5mg/l) for one hour. The DNA fragments were depurinated in 0.25M HCl, transferred onto nylon membrane (Hybond N^+, Amersham) and then hybridized under standard conditions.

In situ hybridization

Recombinant plasmid DNA was biotinylated as described by Langer-Safer (23) with bio-dATP. The probe was then hybridized to fixed squashes of salivary gland cells denatured with 70mM NaOH. Hybridization was overnight at 58°C in 0.6M NaCl, 50mM phosphate buffer (pH: 7.2), 1×Denhardt's solution. The slides were washed three times in 2×SSC at 53°C. The probe was bound to anti-biotin goat IgG and and then to a secondary antibody, peroxidase conjugated rabbit anti-goat IgG. The probe was visualized with o-dianisidine in phase-contrast. Alternatively labelling and fluorescence detection was done with digoxigenin as previously described (24).

RESULTS

Nucleotide sequence of pCp627

The nucleotide sequence of the insert of pCp627 is shown in figure 1. It contains two complete tandem repeat sequences. It is surrounded on both sides of the repeat array by flanking

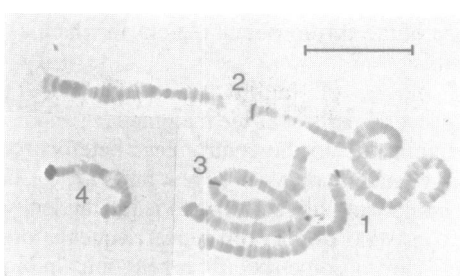

Figure 2. *In situ* hybridization of pCp627 DNA with squashed salivary gland cells of *C. pallidivittatus*. The hybridization signals are detected at the centromeric end of chromosome 4 and at metacentric bands of the three large chromosomes. The bar shows 50μm.

regions. The repeated units have an AT content of 74% and are 96% identical. Each repeat starts with a 37bp region in which two invertrepeats surround a region with 100% AT. The invertrepeats are of two kinds: type A, with the sequence AAA-GCTTT and type B, identical except that the G in the fourth position is duplicated. There is no open reading frame in the insert. Starting from the 5′ end there are 254 positions from the *Eco*RI cloning site to the first repeat. Then follow two repeated units 156 and 154 bp long respectively. After the end of the second repeat there is an A-type invertrepeat. The insert is then terminated with 54 bases that are an inverted version of the beginning of the clone (positions 1- 54).

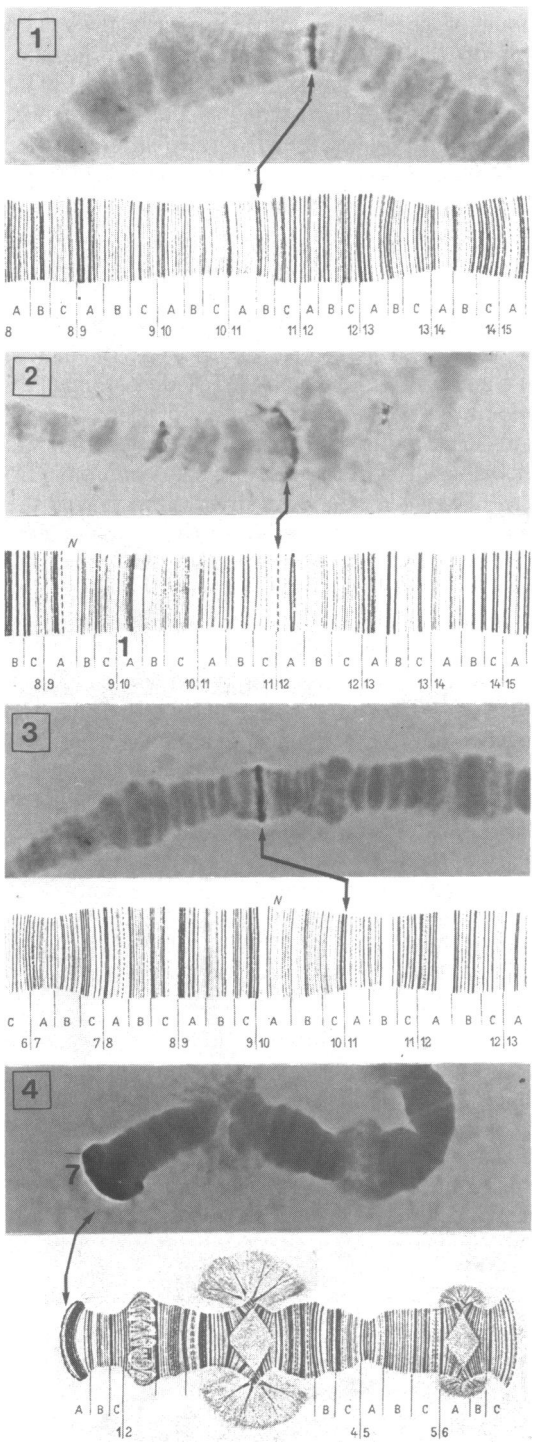

Figure 3. Chromosomal localization of the pCp627 hybridization signal according to the chromosome maps of Beermann (19).

In situ hybridization

The pCp627 cloned DNA was hybridized *in situ* to squashes of salivary gland cells. Label was confined to the centromeric end of chromosome 4 from where the DNA was collected and to putative centromeric bands of the other three metacentric chromosomes (fig 2). Although significant hybridization was detected in these three chromosomes in form of sharp bands, the most intense signal was localized to chromosome 4. Figure 3 illustrates the fine localization of the signal in all four polytene chromosomes. Hybridization was identified in the following positions: for chromosome 1 band 10B1; chromosome 2 band 12A1; chromosome 3 an undetermined band in the 11A region and chromosome 4 the telomeric band 1A1. No hybridization was found elsewhere in the genome. For chromosome 1 the centromere was previously assigned to 11A1. In chromosome 2 the labelling is present in a characteristically dotted band to which it was previously assigned. In all cases the labelling of the metacentric chromosomes falls in central regions free of inversion polymorphisms in agreement with observations that inversions do not encompass the centromere in *Camptochironomus* (17).

The 254 bp piece of DNA flanking the repeats on one side was isolated with *Eco*RI and *Hin*dIII and subcloned in pUC18 (designated pCp254) after which it was hybridized to

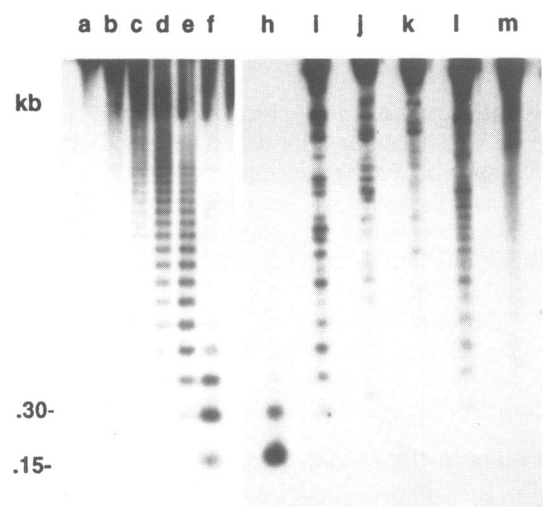

Figure 4. a−f) Partial digestion of total genomic DNA. 5μg of genomic DNA were incubated with 10 units of *Hin*dIII in a 100μl reaction. 10μl aliquots were withdrawn after various times and the reaction stopped by adding EDTA to a final concentration of 10mM. The partially digested genomic DNA was separated in 1% agarose and transferred to nylon filters Hybond N^+ (Amersham) and hybridized with P^{32}-labelled pCp627 DNA. Digestion times were for lanes **a**: 0 min., **b**: 5 min., **c**: 10 min., **d**: 20 min., **e**: 30 min., **f**: 45 min. **h−m)** Digestion of genomic DNA with different restriction endonucleases. In each case 1μg DNA was cleaved and separated in a 1% agarose gel, transferred to nylon filter and hybridized with P^{32}- labelled pCp627 DNA. **h**: *Alu* I, **i**: *Dde* I, **j**: *Hinf* I, **k**: *Rsa* I, **l**: *Sau*3AI, **m**:. *Taq* I

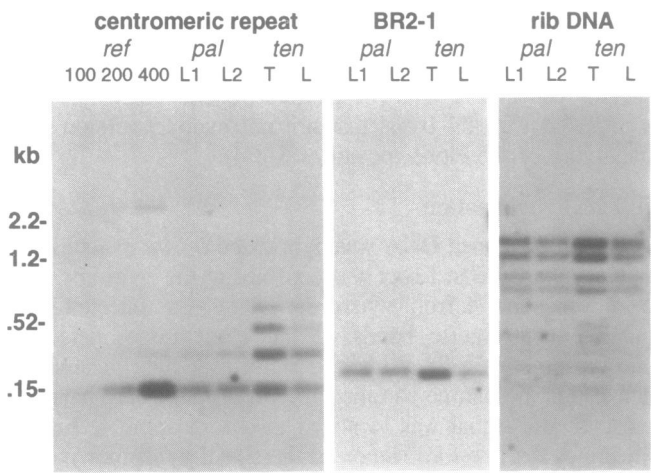

Figure 5. Titration of centromeric 155bp repeat sequences, BR2-1 core repeat DNA and ribosomal DNA (designated *centromeric repeat*, *BR2-1* and *rib DNA*, respectively). As reference (*ref*) was used pCp627 plasmid DNA digested with *Hin*dIII and *Eco*RI in amounts corresponding to 100, 200, and 400kb of 155bp repeat sequences per genome when referred to the quantities of genomic DNA applied. (This is 2.7, 5.35 and 10.7ng of pCp627 DNA when 0.5μg of genomic DNA is used per slot. 0.5μg is 2.5×10^6 genomes, the genome size being 2×10^8 bp or 0.2×10^{-6}mg [ref. 30]. The pCp627 plasmid has a MW=2.0×10^3kD or 3313bp and contains two 155bp repeats, MW=0.19×10^3kD or 310bp. 100kb of 155bp repeat multiplied by 2.5×10^6 is 0.25ng or 2.7ng pCp627 DNA.) Genomic DNA was from *C. pallidivittatus* (*pal*) and *C. tentans* (*ten*). DNA from two different strains of *C. pallidivittatus* larvae was used (*L1* and *L2*) and from one strain of *C. tentans* larvae (*L*) and from tissue cultured cells (*T*), in all cases in the same quantities, 0.5μg. The genomic DNAs were digested with *Hin*dIII and *Hin*fI and separated in 1.2% agarose gels, transferred to HyBond N^+ nylon filters and hybridized with labelled 155bp cloned repeat DNA (left exposure), BR2-1 230bp cloned repeat unit (center exposure) and λCtrDNA, containing a 8.5kb large insert of ribosomal DNA from *C. tentans* (right exposure). The filter is identical in the three exposures and was washed with NaOH between hybridizations.

chromosome squashes and visualized in the fluorescence microscope after labelling with digoxigenin and application of fluorescein-labelled antidigoxigenin. It was necessary to use this method because the signal was too weak in this case for detection with the biotin-peroxidase technique. Signal was detected at the left telomere of chromosome 4 and at the same positions in the metacentric chromosomes as the pCp627 hybridization. In contrast to the labelling with pCp627, representing mainly repeat units, hybridization with pCp254 gave rise to considerable background signal in the form of several dozens of weakly fluorescent chromosome bands (results not shown).

Organization in the genome

In order to reveal the genomic organization of the repeat units we digested genomic DNA with restriction enzymes cutting once in the repeat and with enzymes lacking sites in the two sequenced repeats. The digested DNA was fractionated in agarose, transferred to filter and hybridized with nick-translated pCp627 DNA. Figure 4 (**h**−**m**) shows the results of complete digestion followed by Southern analysis. After application of *Alu*I an approximately 150bp (monomer) as well as a weaker 300bp (dimer) long fragment can be detected. Also enzymes without sites in the two repeated units give fragments corresponding to different multimers. This can probably be explained by base substitutions in the repeat arrays creating occasional sites.

A partial digestion shows that the repeated units in pCp627 are mostly present in long tandem arrays (fig 4 **a**−**f**). A partial digestion with *Eco*RI was done under similar conditions as for *Hin*dIII. We found no evidence for a repeat structure of the whole 627bp unit or for sequences containing the flanking 254bp DNA (data not shown).

Genomic abundance of the pCp627 components

The amount of the 155bp sequence present in the genome of *C. pallidivittatus* was estimated by comparing hybridization signal intensities from restricted genomic DNA with those of the 155bp band from pCp627 in known amounts after electrophoresis and blotting. Larval DNA was used from two different strains of *C. pallidivittatus* and from one strain of the sibling species *C. tentans*. This latter DNA was included because tissue culture cells are available from *C. tentans* and we wanted to compare the amounts in larval, mainly polytenic and nondividing cells with the amounts in actively dividing cells. In a first hybridization the 155bp of the pCp627 insert was used as probe. Other kinds of DNA were used as control probes, viz. Balbiani ring 2-1 (BR2-1) core repeat, 230bp DNA (25) and ribosomal DNA (26). The genomic DNAs were, therefore, digested not only with *Hin*dIII to isolate the 155bp repeat but also with *Hin*fI which cuts once per Balbiani ring repeat and also splits ribosomal DNA into fragments of suitable size. After hybridization with the 155bp insert as probe and exposure of the blot to X-ray film, hybridizing probe was removed and the filter rehybridized with the 230bp BR2-1 plasmid insert. A final hybridization was performed with DNA from a phage containing an 8.5kb *Eco*RI fragment of ribosomal DNA which includes both the 18S and 28S RNA regions. For reference was used pCp627 DNA in amounts that correspond to 100kb, 200kb and 400kb of the 155bp repeat per genome when referred to the quantities of genomic DNA used. It can be seen from figure 5 that the amounts of larval DNA in the samples from the two strains of *C. pallidivittatus* are close to 200kb, i.e. 205 and 210kb, respectively, as determined by photometry of the X-ray film. The amounts of DNA in the two *C. tentans* samples are more difficult to determine. This is because some of the material is present as dimers, trimers and tetramers and the reference is strictly valid only for the monomer. The transfer efficiency during blotting of DNA falls below 200bp and the film blackening for the monomer, therefore, is underrepresented in relation to the di- and multimers. This was made clear by dot blots that were also carried out, in which the amounts of the 155bp DNA were similar in larval DNA from the two species (data not shown). The main interest in the *C. tentans* values lies in the comparison between DNA from larval and tissue cultured cells. In the monomer band there is 300kb for the tissue culture DNA and 195kb for the larval DNA giving a ratio of 1.5. The ratio for all hybridizing material is probably slightly higher since tissue culture DNA contains more di- and multimer material than larval DNA. Filter rehybridized with BR2-1 probe shows similar intensity of hybridization for the three samples of larval DNA. The *C. tentans* signal is slightly weaker which can probably be ascribed to the fact that there is 8% mismatch between the *C. pallidivittatus* probe and the *C. tentans* sequence (25, 27). Again we find distinctly higher level of hybridization for the tissue culture sample, about two-fold. Also ribosomal DNA gave similar excess of hybridization for tissue cultured cell DNA.

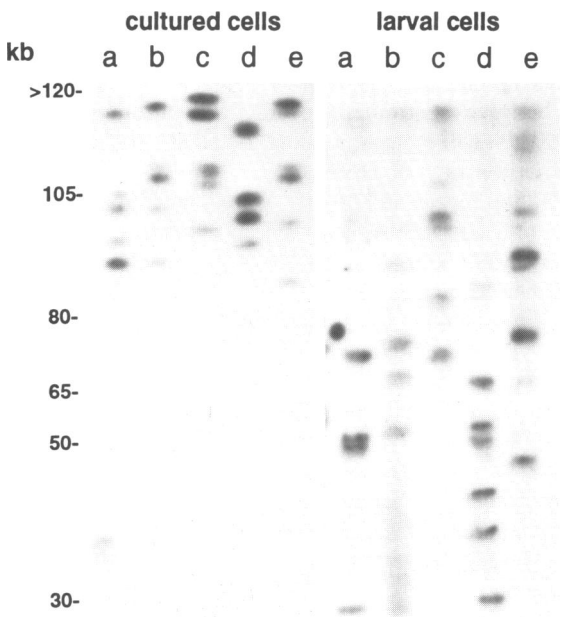

Figure 6. Genomic DNA from tissue cultured cells and larvae was extracted as described in Material and Methods, digested with different restriction endonucleases and resolved by PFGE (GeneNavigator system, Pharmacia-LKB). **a**: *Cla*I, **b**: *Sac*I, **c**: *Sal*I, **d**: *Xba*I, **e**: *Xho*I. After separation the fragmented DNA was depurinated by immersion in 0.25M HCl and transferred to HyBond N^+ nylon membrane. The filter was then probed with P^{32}- labelled pCp155 insert DNA (a plasmid clone derivated from pCp627 that includes only 155bp repetitive units). DNA concatemers of pBr328 partially digested with *Eco*RI (BioRad) were used as molecular size marker ('5kb ladders').

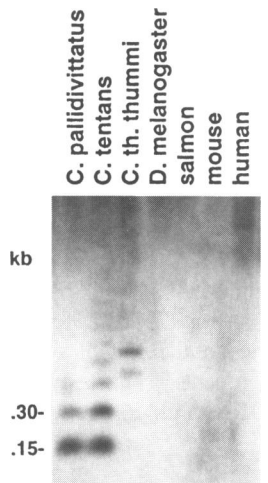

Figure 7. Southern blot analysis in 3% Nu Sieve GTG agarose (FMC BioProducts) of genomic DNA from different species digested with *Alu*I. The filter was hybridized with radiolabelled pCp627 DNA in a solution containing 5×SSPE, 5×Denhardt's solution and 30% formamide for 16hrs at 38°C. The filter was washed twice in 2×SSC: 0.1% SDS at room temperature and once in 2×SSPE: 0.1% SDS at 48°C for one hour.

Since the three kinds of sequences all gave a 1.5−2 fold higher level of hybridization for tissue culture DNA than for larval DNA we controlled the DNA concentrations of the four samples not only by spectrophotometry but also by running samples in 0.3% agarose gel together with known amounts of phage λ DNA for reference. The four samples all contained high molecular weight material similar in migration rate to phage λ DNA and were, therefore, well comparable. The ethidium bromide stained gel was photographed and the negative investigated by photometry. There were only insignificant differences between the different samples of genomic DNA. This result also excludes the alternative that mitochondrial DNA is present to an important extent in larval DNA leading to its dilution, since mitochondrial DNA would migrate as lower molecular weight material, the size of mitochondrial DNA in another dipteran species *Drosophila yakuba* being 16kb (28).

In conclusion we find about 200kb of centromeric DNA sequences per genome in larval DNA. This would correspond to 50kb per centromere, or about 300 repeats if they are evenly distributed between the four centromeres. Judging by the *in situ* results the chromosome 4 centromere contains more repeats. For tissue cultured cell the total amount is 1.5−2 times as high but this also applies to two other kinds of sequences, i.e. BR2-1 DNA and ribosomal DNA. There are two possibilities, i.e. tissue culture cells have acquired a smaller genome and contain the same amounts of 155bp sequence per genome or the genome has not decreased in size which means that also ribosomal DNA and BR2-1 DNA have undergone amplification.

Repeat array sizes as determined by PFGE

High molecular weight genomic DNA was digested with a number of restriction enzymes having six-base recognition sequences, separated by PFGE in 1% agarose together with size markers, blotted and hybridized to labelled 155bp insert. DNA was used from total larval DNA in *C. pallidivittatus* as well as from tissue culture cells from *C. tentans*. Figure 6 shows that bands are obtained in a range of 30 to 90kb for the larval DNA. The variation in pattern is likely to be due to different length of DNA between the repeat array and the relevant restriction site for different enzymes. Another source of variation, possibly more serious, is the occasional occurrence of restriction sites within the arrays due to the relatively frequent mutations in the repeat units. Since a total amount of 200kb is expected for the four centromeres our data suggest that at least the majority of the centromeres contain uninterrupted arrays of tandem repeats. The telocentrically localized centromere in chromosome 4 may, however, be an exception.

The pattern for centromere arrays from tissue culture cells is similar to that of the larval DNA except that it is displaced towards higher molecular weight values. These bands are in the 40kb to more than 120kb range (fig 6), an interesting parallel to the higher content of centromeric DNA in tissue culture cell DNA.

Since the two types of DNA are derived from different species it is possible that some differences exist in restriction patterns. Nevertheless the considerable difference between the sizes of the arrays in the two types of DNA cannot easily be explained by species differences. The closely related, sibling species *C. tentans* and *C. pallidivittatus* contain similar amounts of 155bp repeat DNA in the larval genome. The difference in array sizes provides a parallel to the differences in contents of 155bp repeat DNA between tissue culture cells and larval cells.

Rehybridization of larval DNA blots with pCp254 insert showed that some of the bands containing 155bp repeats also

hybridized with this flanking sequence which also is present in a large number of other bands (results not shown).

Interspecies analysis

Figure 7 shows an autoradiogram of a Southern blot of DNA from different species. The *Alu*I digested DNA was hybridized with labelled pCp627 DNA and washed under permissive conditions. The closely related species *C. pallidivittatus* and *C. tentans* have a similar hybridization pattern. In *C. thummi thummi* signals are seen at about 530 and 700bp. This suggests a slightly larger repeat site than for *C. pallividitattus* and a higher order structure in which most of the *Alu*I sites are distributed at intervals of three or four repeats. There was no detectable cross hybridization with DNA of phylogenetically more remote species.

DISCUSSION

Here we report the isolation and molecular characterization of a tandemly repeated sequence exclusively localized in all the centromeres in *C. pallidivittatus*. Titration experiments indicate that the sequence corresponds to an average of 50kb per centromere. If this is the only repeat present in the centromeres it represents remarkably small amounts of centromeric repetitive DNA, to be compared with quantities in the 500–5000kb range reported for the higher eukaryotes (7). It is, however, in agreement with the fact that *in situ* hybridizations place these repeats in relatively thin bands, a couple of which were earlier assumed to be centromeres, based on cytological observations.

We have no evidence for the presence of any other repetitive element in the centromeres. Furthermore, the thin bands with which the pCp627 hybridizes do not provide space for large quantities of other DNA, the average band size corresponding to 100kb in *Camptochironomus* (29).

Different kinds of repetitive sequences are known to be eliminated during polytenization in *Drosophila*. This probably occurs by a process in which repeats are cleaved out and the resulting free DNA ends rejoined (30). The larval DNA from *Chironomus* is extracted from cells representing all possible degrees of polytenization up to 13 steps (29). If interstitial elimination had taken place this would not be compatible with well defined bands in PFGE.

It is true that we find DNA from actively dividing tissue culture cells containing more centromeric repeats per genome than DNA from larval cells. This applies, however, also to two other kinds of repetitive DNA, a unit from a Balbiani ring gene and ribosomal RNA genes. This phenomenon is surprising and deserves further study. One possibility is that repetitive units, or genes on which there is a demand, become amplified in tissue culture cells, another that there is elimination of other DNA, resulting in a decrease of the genome size. The PFGE results with tissue culture cell DNA suggest that length-wise amplification is an important component since the arrays of centromeric tandem repeats have increased in size roughly in proportion to the increased DNA amounts. The tissue culture line is believed to be of entodermal origin (31). The cells form follicles filled with secretion. It is certainly not excluded that Balbiani ring genes are active and it remains to be investigated if the three sequences studied here are amplified in response to demands for growth, cell division and secretion.

We find that the repeat array is flanked on one side by a unique fragment of 254bp and on the other by a minor part of this fragment in an inverted arrangement. The *in situ* hybridizations show that the flanking element is present in all of the centromeres. It is likely that it can be used for extending the sequence characterization outwards from the repeat array and to determine limits between centromere-specific and chromosome-specific DNA. This will permit a more detailed characterization of centromeric arrays of DNA.

The 155bp repeats are present as long tandem arrays whereas pCp627 only contains two such repeats in tandem. One possibility is that arrays of a larger size were originally cloned and several repeats then eliminated. There is, nevertheless, no possibility that anything substantially larger than 7kb, the upper cloning capacity of the λ vector, was cloned. It is also possible, as future work will have to decide, that the telocentrically localized centromere is organized differently than the interstitial centromeres.

The repeat unit has a remarkable structure containing at one end a pair of invertrepeats with similar sequence in a very AT-rich enviroment. Whereas the proximal invertrepeats form a perfect palindrome, the central part of which constitutes a *Hin*dIII site, the distal invertrepeats have the central G duplicated, giving the sequence AAAGGCTTT. The arrangement shows some similarity to the *Saccharomyces* centomere where a central functionally important element is rich in AT and adjoins regions with sequences showing dyad symmetry. The pCp627 clone shows striking variations in GC content with up to 60% in the region flanking the repeats down to zero in the DNA connecting the two invertrepeats. The central parts of the repeats show more typical levels of GC content for *Chironomus*, i.e about 30%. An AT-rich DNA is of interest when localized close to invertrepeats since it may promote transitions from linear to cruciform DNA conformations (32).

If the repeat unit characterized here is critically involved in centromeric function one could perhaps expect evolutionary sequence conservation at least in regions interacting with proteins. The α-satellite found in centromeric regions of higher eukaryotes is, however, rapidly evolving but it has been pointed out that this does not necessarily argue against a functional role (8). The evolutionary constancy of a series of repetitive families has been studied in *Chironomus*. In all of the cases their sequence evolution is rapid and they can only be traced between closely related species, in no case as far as between the distantly related *C. thummi thummi* and *C. pallidivittatus* (33). It is therefore a significant result when we find that pCp627 cross-hybridizes to *C. thummi thummi*. Nevertheless the sequence constancy is not very high in part of the repeat units. This is shown, firstly by genomic blots in which a number of restriction enzymes with no sites in any of the two sequenced units have frequent cutting sites in the genomic repeats, secondly by the sequence comparison between the two repeats.

ACKNOWLEDGEMENTS

This investigation was supported by grants from the Swedish Cancer Society and the Philip-Sörensen Foundation. We are indebted to Dr Deidre Gilligan, ETH, Zürich for a gift of DNA from *C. tentans* tissue cultured cells, to Dr Erwin R. Schmidt, Mainz, for *C. thummi* larval DNA and to Dr Steven T. Case, Jackson, Mississippi, for the λCtr DNA clone.

REFERENCES

1. Clarke, L. and Carbon, J. (1980) *Nature* **287**, 504–509.
2. Clarke, L., Amstutz, H., Fishel, B. and Carbon, J. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8253- 8257.

3. Nakasenko, Y., Adachi, Y., Funahashi, S. Niwa, O. and Yanagida, M. (1986) *EMBO J.* 5, 1011–1021.
4. Clarke, L. and Baum, M.P. (1990) *Mol. Cell. Biol.* **10**, 1863–1872.
5. Miklos, G.L. (1985) In McIntyre, R.J. (ed.), Molecular Evolutionary Genetics. Plenum Press, New York, pp. 241–321.
6. Pardue, M.L. and Gall, J.G. (1970) *Science* **168**, 1356–1358.
7. Willard, H. and Waye, J.S. (1987) *TIG* **3**, 192–198.
8. Willard, H.F. (1990) *TIG* **3**, 192–198.
9. Grady, D., Ratliff, R., Robinson, D., McCanlies, E., Meyne, J. and Moyzis, R. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 1695–1699.
10. Wong, A.K.C. and Rattner, J.B. (1988) *Nucleic Acids Res.* **16**, 11645–11661.
11. Joseph, A. Mitchell, A.R. and Miller, O.J. (1989) *Exp. Cell Res.*183, 494–500.
12. Earnshaw, W.C. and Rothfield, N. (1985) *Chromosoma* 91, 313–321.
13. Masumoto, H., Masukata, H., Muro, Y., Nosaki, N. and Okasaki, T. (1989) *J. Cell. Biol.* **109**, 1963–1973.
14. Haaf, T., Warburton, P. and Willard, H. (1992) *Cell* 70, 681–696.
15. Gall, J. G. and Atherton, D.D. (1974) *J. Mol. Biol.* **85**, 633–634.
16. Schmidt, E. R. (1981) *FEBS letters*, **129**, 21–24.
17. Beermann, W. (1955) *Chromosoma* **7**, 198–259.
18. Scalenghe, F., Turco, E., Edström, J-E., Pirrotta, V. and Melli, M. (1981) *Chromosoma* **82**, 201–216.
19. Beermann, W. (1952) *Chromosoma* 5, 139–198.
20. Shen, C.K. and Maniatis, T. (1980) *Cell* **19**, 379–391.
21. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning, A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press.
22. Bingham, P. M., Levis, R. and Rubin, G.M. (1981) *Cell* 25, 693–704.
23. Langer-Safer, P.R., Levine, M. and Ward, D.C. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4381–4385.
24. Cohn, M. and Edström, J-E. (1992) *Chromosoma* 101, 634–640.
25. Galler, R., Saiga, H., Widmer, R.M., Lezzi, M. and Edström, J-E. (1985) *EMBO J.* 4, 2977–2982.
26. Dreesen, T.D., Bower, J.R. and Case, S.T. (1985) *J. Biol. Chem.* 260, 11824–11830.
27. Sümegi, J.L., Wieslander, L. and Daneholt, B. (1982) *Cell* **30**, 579–587.
28. Clary, D.O. and Wolstenholme, D.R. (1985) *J. Mol. Evol.* 22, 252–271.
29. Daneholt, B. and Edström, J-E. (1967) *Cytogenetics* **6**, 350–356.
30. Glaser, R.L., Karpen, G.H. and Spradling, A.C. (1992) *Chromosoma* 102, 15–19.
31. Wyss, C. (1982) *Exp. Cell. Res.* **139**, 309–319.
32. Sullivan, K.M. and Lilley, D.M.J. (1986) *Cell* **47**, 817–827.
33. Schmidt, E. R. (1984) *J. Mol. Biol.* **178**, 1–15.