

# Rapid generation of long synthetic tandem repeats and its application for analysis in human artificial chromosome formation

Tom Ebersole, Yasuhide Okamoto<sup>1</sup>, Vladimir N. Noskov, Natalay Kouprina, Jung-Hyun Kim, Sun-Hee Leem, J. Carl Barrett, Hiroshi Masumoto and Vladimir Larionov\*

Laboratory of Biosystems and Cancer, National Cancer Institute, Bethesda, MD, USA and <sup>1</sup>Division of Biological Science, Graduate School of Science, Nagoya University, Chikusa-ku, Nagoya 464-8602, Japan

Received June 15, 2005; Revised and Accepted August 3, 2005

## ABSTRACT

**Human artificial chromosomes (HACs) provide a unique opportunity to study kinetochore formation and to develop a new generation of vectors with potential in gene therapy. An investigation into the structural and the functional relationship in centromeric tandem repeats in HACs requires the ability to manipulate repeat substructure efficiently. We describe here a new method to rapidly amplify human alloid tandem repeats of a few hundred base pairs into long DNA arrays up to 120 kb. The method includes rolling-circle amplification (RCA) of repeats *in vitro* and assembly of the RCA products by *in vivo* recombination in yeast. The synthetic arrays are competent in HAC formation when transformed into human cells. As short multimers can be easily modified before amplification, this new technique can identify repeat monomer regions critical for kinetochore seeding. The method may have more general application in elucidating the role of other tandem repeats in chromosome organization and dynamics.**

## INTRODUCTION

Tandem repeat arrays are present throughout the genomes of eukaryotes and play important roles in creating and maintaining specialized chromatin (e.g. at centromeres and telomeres) and are often associated with heterochromatin (1,2). Small tandem repeat arrays also play a role in gene regulation (3–5), and variants have been linked to human disease or disease likelihood (6–9). They also may play a role in rapid evolution (10,11). Centromeric tandem repeats are associated

with the functional kinetochore, the structure that attaches to spindle microtubules for chromosome partitioning to daughter cells. The centromeres of most of the higher eukaryotes that have been studied so far contain tandem repeat arrays of hundreds to thousands of kilobases in size, including centromeres of plants, invertebrates and vertebrates (12–14).

Alloid arrays at human centromeres extend over many millions of base pairs. Type I arrays are composed of highly homogeneous higher-order repeats (HOR) of the 170 bp monomer that are unique to a specific chromosome or shared by a few chromosomes (1). Type I arrays are believed to be an important DNA component of a functional centromere for at least two reasons. First, type I arrays associate with centromere proteins such as CENP-A, which closely interact with DNA to form the kinetochore (15,16). Second, type I arrays are competent to form human artificial chromosomes (HACs) when transformed into human cells (17–25).

HACs represent extra chromosomes carrying all the required components of a functional kinetochore. HACs have advantages as gene expression vectors with potential for use in gene therapy. They are stably maintained at a low copy number in the host nucleus. They should not cause the problem of severe immunogenic responses associated with adenoviral vectors because HACs contain no viral genes or proteins. They are suitable for carrying intact mammalian genes that should confer physiological levels of fully regulated gene expression because the HACs are surrounded by all of their long range controlling elements. Several groups have had success in complementing a genetic deficiency with HACs carrying the full-size gene (20–22).

Early HAC formation studies used only a few of the many subfamilies of alloid DNA arrays that were identified in bacterial artificial chromosome (BAC) and yeast artificial chromosome (YAC) libraries. Alloid arrays with monomers containing the 17 bp CENP-B box from chromosomes 21, X, 17 and 5 cloned into YAC, BAC or phage artificial

\*To whom correspondence should be addressed. Tel: +1 301 496 7941; Fax: +1 301 480 2772; Email: larionov@mail.nih.gov

chromosome (PAC) vectors have been shown to be competent to form *de novo* artificial chromosomes in cultured cells, whereas arrays lacking the CENP-B box from the Y chromosome, chromosome 21 type II and chromosome 22 have proven to be inefficient (17–25). Recently, the requirement for the CENP-B box for *de novo* centromere and HAC assembly was demonstrated using synthetic type I alphoid DNAs containing functional CENP-B boxes or mutant CENP-B boxes (24–26). However the presence of the CENP-B box is not sufficient to predict an effective array. X chromosome arrays that contain CENP-B boxes are relatively poor substrates when compared with chromosome 17 derived arrays (27). Substitution of alphoid sequence outside the CENP-B box by GC rich DNA in a synthetically constructed array demonstrated that the CENP-B box alone is not sufficient for centromere nucleation (24). The CENP-B box consists of 17 bases, a subset of which constitutes the core residues required for efficient CENP-B binding (28–30). Apart from the identification of these core residues, the bases of the alphoid monomer that are essential for successful centromere nucleation remain unknown. AT richness is found in the centromere repeats of many organisms, including human alphoid repeats, but it is yet to be determined whether the abundance of these two bases is somehow related to function or if particular bases are critical.

Large alphoid tandem repeat DNA segments isolated from genomic libraries are difficult to fully characterize and cannot be modified readily. Therefore, further analysis of alphoid DNA arrays with a defined sequence is required to elucidate the structural requirements for efficient *de novo* assembly of centromere structure. To accelerate such analysis, we developed a novel strategy to rapidly construct synthetic alphoid DNA arrays with a predetermined structure. This technique is comprised of two steps: rolling-circle amplification (RCA) of a short alphoid DNA multimer (e.g. a dimer) and subsequent assembly of the amplified fragments by *in vivo* homologous recombination in yeast. Using this method, we constructed a set of different synthetic alphoid DNA arrays varying in size from 30 to 120 kb and demonstrated that they can be competent in HAC formation. As any nucleotide can be easily changed in an alphoid dimer before its amplification, this new technique is optimal for identifying the critical regions of the alphoid repeat for *de novo* centromere seeding. Practicable manipulation of alphoid or other types of repeats can also be a basis for experiments to elucidate the critical substructure that leads to heterochromatin formation.

## MATERIALS AND METHODS

### Rolling-circle amplification

RCA was performed using an Amersham TempliPhi kit according to the manufacturer's instructions, except that reactions were scaled up to 100  $\mu$ l and were spiked with a template-specific primer mix to a final concentration of 2 pmol/ $\mu$ l. RCA primers for alphoid DNA were 5'-AATCTGCA-3', 5'-ACTAGACA-3' and 5'-ACAGAGTT-3' for the upper strand, and 5'-AGAGTGTT-3', 5'-TCTGAGAA-3' and 5'-GGCCTCAA-3' for the lower strand. Primers for mouse major satellite were 5'-ACTTGACGA-3' and 5'-TGCACTGA-3' for the upper strand, and 5'-TTAGAAATGT-3' and 5'-GAATATGGCG-3' for the lower

strand. Primers for mouse minor satellite were 5'-AATGAGTT-3' and 5'-TTCGTTGGAAACGGG-3' for the upper strand, and 5'-AGTGTGGTT-3' for the lower strand. Primers for human gamma-8 satellite were 5'-AATTCTGGG-3' for the upper strand, and 5'-CCAGAATT-3' and 5'-GACACCTC-3' for the lower strand. Primers for the human *Alu* repeat were 5'-AATGTAGC-3' and 5'-TCCTGAGCTCA-3' for the upper strand, and 5'-GTAATCCC-3' for the lower strand. All RCA primers carried thio-modified phosphate linkages for the last two bases of the 3' end. Target templates were obtained by PCR from genomic DNA for mouse major, mouse minor and human gamma-8 satellites, and by PCR from a cloned human *HPRT* gene for the *Alu* repeat. Primers contained a restriction enzyme site such that the circular template would reconstitute a complete monomer after ligation. Typically one or two bases were substituted at the ligation junction as a result of the introduced restriction enzyme site. PCR primers are summarized in Supplementary Table 1S. PCR products were cloned into Invitrogen Topo vectors. The 2mer, 4mer and 5mer human alphoid template DNAs were obtained by cloning directly into the pBluescript II EcoRI site from an EcoRI digested PAC clone containing ~35 copies of the human chromosome 21 11mer. The accession number of the complete sequence of the alphoid 11mer is GI:550080. The positions of the 2mer sequence in the 11mer are 850–1193. This sequence was modified to obtain dimers with a different level of homology between the monomers. In one of the modified sequences, six nucleotide substitutions were done in one monomer (6S type dimer). In another dimer, a 40 bp sequence of one of the monomers (positions 987–1026 in GI:55080) was replaced by a heterologous 40 bp sequence (Het type dimer).

Circular reaction templates were generated from gel-purified and ligated inserts derived from clones in pUC-base plasmids. Ligation was performed under dilute conditions at ~1 ng/ $\mu$ l. Circular templates were directly mixed into the RCA reaction at ~0.1–0.2 ng per 10  $\mu$ l of reaction. Reaction products were phenol/chloroform extracted and ethanol precipitated prior to cloning. The size range and quantity of output double-stranded DNA is similar to that of a control reaction using pUC19 and random hexamers.

### Extension of RCA products by recombinational cloning in yeast

RCA products were cloned in yeast using a vector with appropriate hooks. Size of alphoid satellite hooks was ~40 bp. For other types of repeats the size of the hooks was ~100 bp (Supplementary Table 2S). The basic targeting vector TAR-NV contains YAC (*HIS3*, *CEN6*, *ARSH4*) and BAC (*Cm*, *ori F*) cassettes as well as a mammalian selectable marker (*Neo* or *BS*). Before transformation the vectors were linearized to release targeting hooks. The highly transformable *Saccharomyces cerevisiae* strain VL6-48N (*MATalpha*, *his3- $\Delta$ 200*, *trp1- $\Delta$ 1*, *ura3- $\Delta$ 1*, *lys2*, *ade2-101*, *met14*), which has *HIS3* and *URA3* deletions, was used for transformation. Conditions for spheroplast transformation were described previously (31). Each transformation used 2–3  $\mu$ g of RCA product and 0.02  $\mu$ g of the linearized vector. Typically, at such conditions, 200–1000 transformants were obtained. Omitting the RCA product from the transformation mix resulted in a

decreased yield of transformants, down to 5–20 colonies. Individual His<sup>+</sup> transformants were streaked onto SD-His plates (~100 colonies per plate), incubated overnight at 30°C and used for isolating high molecular weight yeast DNA. In the first experiments we determined the size of inserts in individual transformants by blot hybridization; however, in subsequent experiments we combined 200–300 YAC clones and purified genomic DNA, then electroporated the DNA into *Escherichia coli* cells (DH10B or Stb14; Invitrogen) to obtain repeat arrays directly in a BAC form. Inserts were sized by CHEF after NotI digestion of BAC DNA isolated from at least 40 bacterial transformants for each construct. In some cases when yield of yeast transformants was low, a second round of recombinational cloning was carried out to increase the size of inserts. For this purpose, 5 µg of BAC DNA with the largest insert was digested with Sall to cleave it at insert/vector junctions. The vector DNA was eliminated with an additional Sau3AI digestion. The final digest was precipitated with ethanol/sodium acetate and dissolved in 20 µl of water. Yeast spheroplast transformation used 3–4 µg of digested DNA and 0.2–0.3 µg of the linearized vector. The yield of clones with 2- to 3-fold larger insert size was 2–5%.

#### Cell culture and BAC DNA transfection

Human fibrosarcoma cell line HT1080 was grown in DMEM medium supplemented with 10% fetal bovine serum (Invitrogen), penicillin, streptomycin and glutamine. BAC DNA (400 ng) was purified using a Qiagen Large Construction kit (Qiagen) and transfected into  $6 \times 10^5$  HT1080 cells

using Lipfectamine reagent (Invitrogen) according to the manufacture's instructions. Stable transformants were selected with 400 µg/ml G418 (Wako).

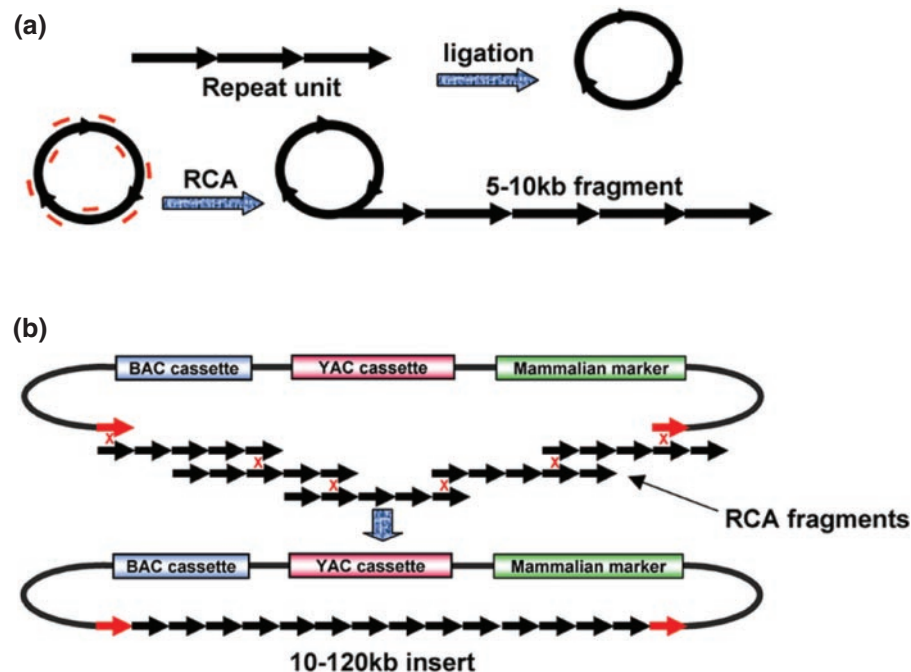
#### Cytological detection of HACs

Standard techniques for fluorescence *in situ* hybridization (FISH) were carried out for the alphoid BAC transformed cell lines as described previously (32). Probes were as follows: p11-4 alphoid DNA for the 5mer alphoid DNA; PCR products amplified from pBAC108L using primers BACX and BACS for BAC vector DNA; and PCR products amplified from HT1080 genome using three sets of primers [a(1)18a/b, a(Y)a/b and CB15a/b] for pan-alphoid DNA as described in our previous reports (24,33,34). Plasmid DNA or PCR products were labeled using a nick translation kit with digoxigenin-11dUTP or biotin16-dUTP (Roche Diagnostics). Indirect immunofluorescence and simultaneous staining by FISH were carried out as described previously (33). Antibodies used were anti-CENP-A [mAN1 (34)], anti-CENP-B [2D8D8 (22)] and anti-CENP-E [mAb177 (35)]. Images were captured using a cooled-CCD camera (PXL; Photometrics Ltd) mounted on a Zeiss microscope and analyzed by IPLab software (Signal Analytics).

## RESULTS

### Construction of synthetic tandem arrays

The first step in the generation of synthetic tandem arrays involves *in vitro* RCA of repeats (Figure 1a). Phi 29



**Figure 1.** Schematic representation of construction of synthetic tandem arrays. (a) The first step includes amplification of multimers by RCA to 5–10 kb. Repeat-specific exonuclease-resistant primers are used for an efficient RCA reaction. (b) The second step includes co-transformation of the RCA-amplified fragments into yeast cells along with a vector containing alphoid-specific hooks. End-to-end recombination of alphoid DNA fragments followed by interaction of the recombined fragments with the vector results in the rescue of large arrays as circular YACs in yeast is shown. The vector contains an yeast cassette, *HIS3/CEN/ARS* (a selectable marker *HIS3*, a centromere sequence *CEN6* from yeast chromosome VI and yeast origin of replication *ARSH4*) and a mammalian selectable marker (the *Neo* or *BS* gene) and a BAC replicon that allows the YAC clones to be transferred into *E.coli* cells.



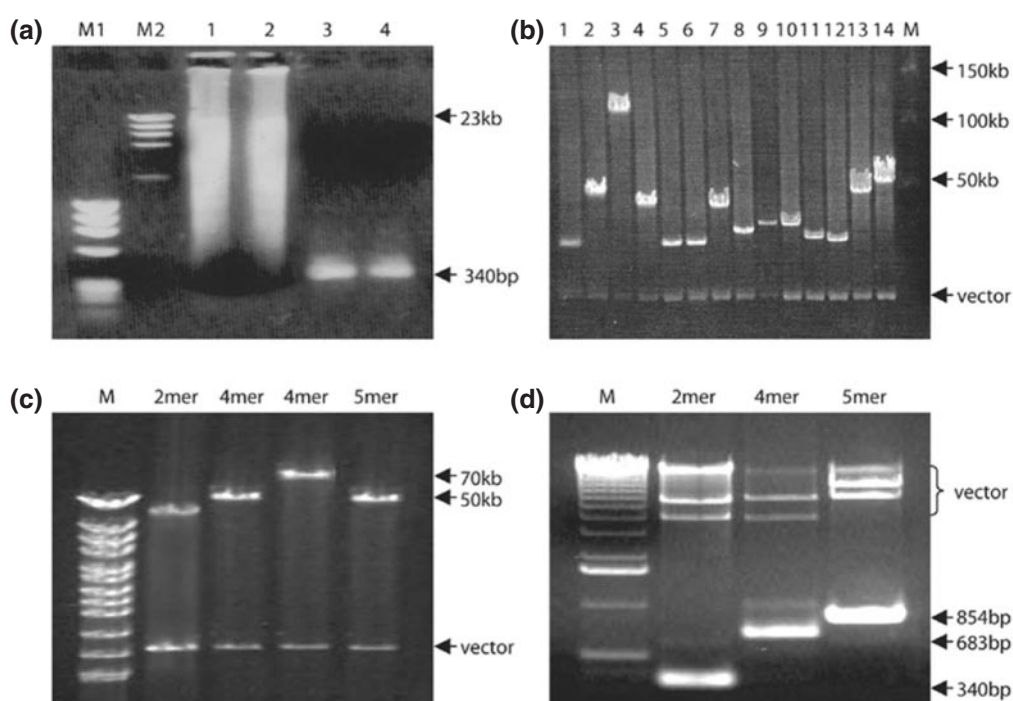
polymerase has a high processivity and can extend newly replicated strands from circular double-stranded templates for several kilobases *in vitro*. Multiply-primed RCA results in hyper-branching of newly synthesized strands yielding exponential amplification in copy number. Priming of 'hyper-branched' RCA is routinely achieved with random hexamers on complex DNA (36,37). The low complexity of tandem repeat DNA, however, results in inefficient amplification with random primers. Therefore, for alphoid DNA repeats as well as for other types of repeats, specific exonuclease-resistant primers based on conserved regions of the repeat monomer were synthesized (see Materials and Methods section). As template DNA, cloned fragments derived from BAC inserts or PCR products amplified from genomic DNA were gel purified and formed into circles by ligation. Cleavage and primer sites were chosen to reform a complete monomer upon ligation. Starting circular template taken from a dilute ligation reaction was as low as 0.1 ng per 10  $\mu$ l of RCA reaction.

Dimer, 4mer and 5mer repeats of the alphoid 170 bp monomer were first used for RCA. All of these are derived from the human chromosome 21 type I 11mer HOR (Supplementary Figure 1S) (18,24). The smallest template DNA used here was the double-stranded 340 bp alphoid dimer. Figure 2a illustrates RCA reactions for a 340 bp alphoid DNA dimer. Although bands with mobility higher than 20 kb are seen, they are likely to be multibranched DNA molecules (37) having anomalous migration (Figure 2a, lanes 1 and 2). Cleavage of reaction products with an appropriate enzyme results in the restoration of the input template fragment (Figure 2a, lanes 3

and 4), demonstrating the faithfulness of the polymerization. Similar results were obtained for RCA reactions with the 4mer, the 5mer and the 6mer (data not shown). The DNA yield from a 100  $\mu$ l multiply-primed RCA reaction is sufficient for several yeast transformation experiments to further expand the size of repeats.

The second step involves assembling of the RCA products into long alphoid DNA arrays by *in vivo* homologous recombination in yeast. For this purpose, the RCA amplified products are co-transformed into yeast spheroplasts along with the targeting vector TAR-NV (Figure 1b). Homologous recombination between the ends of RCA products results in the rescue of large tandem arrays in the targeting vector as circular YACs. By the use of a mixture containing 0.02  $\mu$ g of the targeting vector and 3  $\mu$ g of RCA reaction product generated from alphoid DNA units, between 200 and 1000 His<sup>+</sup> transformants were typically obtained.

The first experiments on amplification by homologous recombination in yeast were carried out with the 5mer-based RCA concatamers. CHEF analysis of the YAC clones demonstrated that 20% of yeast transformants (24/120) contained alphoid DNA inserts larger than 15 kb. Five percent of the transformants contained YACs with a size that varied from 30 to 140 kb. To confirm that the large size of arrays resulted from recombinational interaction between RCA products but not from the capture of pre-existing large DNA molecules generated by RCA, we isolated a 5 kb array from the YAC clones and re-transformed it along with the TAR-NV vector into yeast. Analysis of 40 randomly selected transformants



**Figure 2.** Generation of large alphoid arrays. (a) Multiply-primed RCA reaction products from a 340 bp alphoid dimer (lanes 1 and 2) that retain tandem repeat structure as shown by EcoRI restriction enzyme digestion (lanes 3 and 4). (b) The YAC/BACs generated from the 5mer-based RCA product by recombinational cloning. Fourteen randomly picked BACs obtained after transferring a pool of clones into *E.coli* cells are shown. The size of inserts varies from 30 to 120 kb. (c) Array size for alphoid 2mer, 4mer and 5mer. (d) Origin of insert arrays is confirmed by EcoRI digestion. The upper bands represent vector fragments. The 5mer-based array differs from 2mer- and 4mer-based arrays because this array was assembled using the TAR-NV vector variant that lacked a BAC cassette. The YAC clone was then converted into YAC/BAC with the BRV1 retrofitting vector (25).

**Table 1.** Synthetic arrays generated from different types of repeats

Repeat unit	Size of the unit (in kb)	Size of arrays (in kb)	Fold increase
Human alphoid DNA			
2mer	0.34	30, 35, 40	×118
2mer <sup>a</sup>	0.34	30, 35, 50	×147
2mer <sup>b</sup>	0.34	35, 40, 60	×176
4mer	0.68	40, 50, 70	×103
5mer	0.85	50, 120, 140	×165
6mer	1.02	35	×35
Mouse satellite DNA			
Mouse major satellite, 3mer	0.7	25, 40, 55	×79
Mouse minor satellite, 4mer	0.5	10 <sup>c</sup>	×20

<sup>a</sup>In this 2mer 6 nucleotides are mutated (6S type dimer).

<sup>b</sup>In this 2mer a 40 bp sequence in one of the monomers is replaced by heterogeneous sequence (a Het type).

<sup>c</sup>Apparent small size of inserts is due to the limited number of transformants analyzed.

revealed seven clones containing YACs larger than 20 kb (data not shown). Thus an efficient end to end recombination of incoming DNA molecules during yeast transformation results in a recovery of clones with large alphoid arrays.

Alphoid DNA arrays generated from the 5mer were efficiently and accurately transferred into *E.coli* (Figure 2b). As alphoid DNA arrays are reasonably stable in *E.coli* cells, analysis of the assembly of other types of repeats was carried out by omitting size determination of yeast inserts. The yield of large inserts with 2mer, 4mer and 6mer-based arrays was determined directly in bacterial cells after transferring the yeast isolates into *E.coli*. For this purpose, a pool of the primary YAC clones was electroporated into *E.coli* and size of BACs in individual transformants was determined (see Material and Methods section). Between 5 and 7% of the *E.coli* transformants contained BACs with insert size varying between 20 and 70 kb (Table 1). Thus, combination of RCA with a recombinational capture in yeast may increase the original size of a repeat up to 176 times. Several alphoid DNA clones generated by *in vivo* recombinational cloning are shown in Figure 2c and d. Random sequencing from cloned arrays indicates that the resulting arrays faithfully reflect input template DNA (data not shown). Non-alphoid tandem arrays were also synthesized, including those composed of mouse major and minor satellite, human gamma-8 satellite and human *Alu* repeat. These arrays were then cloned by recombination in yeast using targeting vectors with appropriate hooks (Table 1 and Table 1S).

We conclude that *in vivo* recombination in yeast is highly efficient in assembling the fragments containing tandem repeats into large DNA arrays.

### Stability of synthetic centromeric tandem repeat inserts

The synthetic arrays generated by RCA and recombinational cloning have a higher sequence identity per unit length than their endogenous counterparts, and therefore, may be less stable when cloned. However, 40–120 kb arrays generated from the 4mer and 5mer did not show significant instability in yeast. Clones containing alphoid DNA fragments isolated from chromosome 21 (11mer-based array) and the clones with synthetic arrays derived from the 4mer or 5mer revealed single

bands after their linearization followed by Southern blot hybridization (data not shown). More important was that these inserts were also reasonably stable structurally during their propagation in a *recA* bacterial host (DH10B) at 30°C (Figure 3a, b and d). Growth of the cells at higher temperature (37°C) resulted in some structural instability in the large blocks of alphoid DNA.

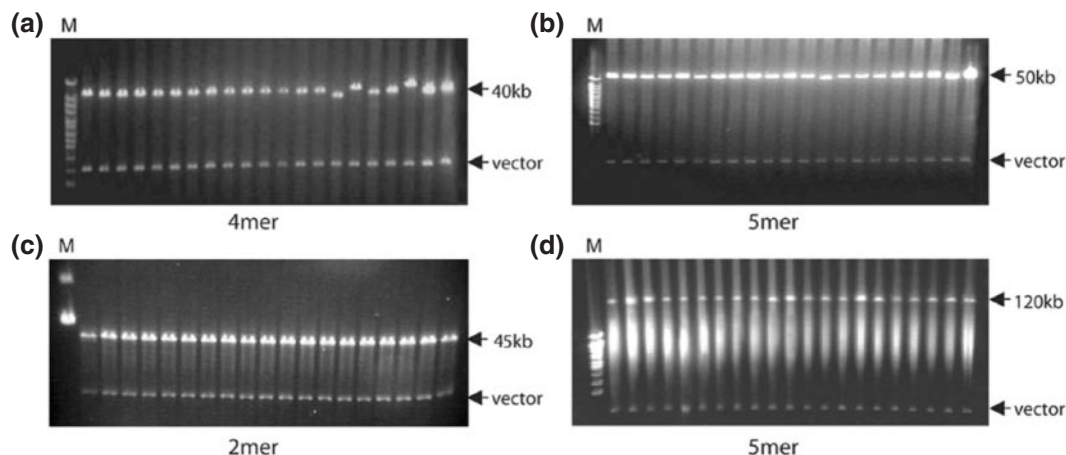
We also analyzed the structural stability of 2mer-based arrays with a different level of homology between the monomers. One dimer corresponds to the chromosome 21 alphoid DNA sequence (accession number GI:550080) designated as a CENP-B-plus type. Two others were mutagenized before amplification. Specifically, we replaced six nucleotides in the dimer (6S type dimer) and replaced a part of one of the monomers with a 40 bp heterologous sequence (a Het type) (see Material and Methods section). These changes reduced homology between monomers in the dimers from 78 to 75 and 61%, respectively. Similar to the 4mer and 5mer-based arrays, 40–60 kb arrays generated from the 6S and Het types of 2mer-based RCA products were structurally stable in a BAC vector. No detectable rearrangements were observed in the subclones during propagation at 30°C (Figure 3c). A similar stability was observed for the CENP-B type arrays with size up to 30 kb. However, clones of this type of array were structurally unstable when the size of inserts exceeded 35 kb. Small deletions were observed in 10–20% of subclones (data not shown). Such instability may be due to a lower level of divergence between monomers in the dimer compared with that of the 6S type and Het dimers.

We conclude that, in general, large size synthetic arrays generated from alphoid DNA repeat units are reasonably stable in *recA* hosts and therefore, they can be used as a substrate for HAC formation.

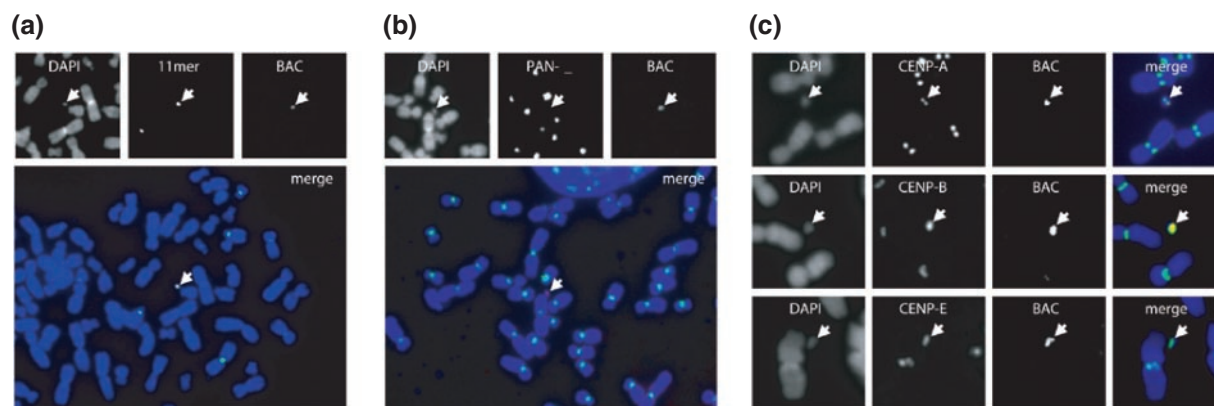
### A synthetic alphoid DNA array in HAC formation

All HACs reported to date have used a native HOR as the basic repeat structure for the centromeric sequence. It is not known if artificially constructed arrays are competent for *de novo* centromere formation in human cells. To further validate the cloned arrays, we attempted to generate HACs in cultured cells using the ~120 kb 5mer-based synthetic array. The 5mer array was derived as a subfragment of the human chromosome 21 11mer HOR that has been used successfully for *de novo* HAC formation (24). The 5mer array contains a CENP-B box density similar to that of the 11mer (2.63 and 2.35 per kb, respectively). The native 11mer contains one monomer with a mutant CENP-B box that cannot bind CENP-B (24). The 5mer retains this monomer. The ratio of mutant to canonical CENP-B boxes is elevated 3.4-fold in the 5mer.

Following lipofection of BAC DNA to HT1080 cells and G418 selection, 29 resistant cell lines were expanded and examined for the presence of HACs by dual FISH with BAC and human chromosome 21 alphoid probes. Three cell lines (10%) contained candidate HACs with 50% or more of individual mitotic cell spreads showing HAC signals (Figure 4a). A control transfection done in parallel using a BAC with a 60 kb insert of the complete 11mer yielded 17% of examined colonies with HACs in at least 50% of cells (data not shown). Size and copy number of the HACs were in the range normally reported for *de novo* formation. A pan-alphoid



**Figure 3.** Stability of synthetic 2mer, 4mer and 5mer based alphoid arrays. To analyze the stability of the alphoid arrays, transformants were streaked to single colonies, and individual subclones were analyzed by CHEF. Of 19–21 independent *E. coli* subclones for each construct, only a few showed a different insert size due to deletions/rearrangements. (a) 4mer, (b) 5mer, (c) 2mer and (d) 5mer.



**Figure 4.** HAC formation using the 120 kb synthetic alphoid 5mer-based array. (a) Both chromosome 21 specific alphoid and BAC vector probe detect the HAC (arrows). Additional signal in the alphoid probe and merged panel detect endogenous chromosome 21 centromere in HT1080 cells. (b) Validation of the HAC in the clone HT4-10. The pan-alphoid probe (blocked for chromosome 21 alphoid) does not detect the HAC. (c) Detection of HACs with anti-CENP-A, -B and -E antibodies.

probe (blocked for chromosome 21 specific alphoid) did not hybridize to the HACs (clone HT4-10 in Figure 4b), suggesting that these three HACs have been assembled without recruiting any endogenous, functioning centromere sequences. Additional evidence that the 5mer array formed the functional centromere *de novo* was binding the candidate HACs to CENP-A and CENP-E, two centromere proteins found at functioning kinetochores, as well as observation of strong CENP-B signal on the 5mer (Figure 4c).

## DISCUSSION

Relatively rapid construction of defined alphoid construct variants will greatly facilitate exploration of the sequence requirements for *de novo* centromere assembly. Previously, two groups reported the construction of synthetic alphoid arrays using repetitive directional ligation on the basis of a native higher-order 2–3 kb repeat fragment (17,24,26). Although this approach allows construction of large synthetic arrays, it has some limitations. First, it is a slow, laborious strategy not

easily scaled up for rapid generation of tandem repeats with engineered changes. In addition, the method tends to result in the use of restriction sites that may not be available in an amplified unit. In contrast, artificially introduced restriction sites remain in multiple copies in the final constructs.

In this work we describe a new strategy to generate large synthetic DNA repeats with a predetermined structure by *in vivo* recombination in yeast. The method includes concatamerization of DNA into short repeats (using RCA or directional *in vitro* ligation), followed by assembly of the short repeats into long arrays by homologous recombination during transformation into yeast cells. Using this system, we generated large synthetic arrays from the different ‘units’ of alphoid DNA. Up to 120 kb arrays generated from the 4mer and 5mer were stable during their propagation as BACs in *E. coli* cells, similar to native HOR alphoid DNA fragments isolated from chromosome 21. Surprisingly, structurally stable alphoid DNA 2mer-based arrays can also be generated when monomers in the dimer are >25% diverged. Size of the constructed arrays exceeds the minimal size that is required for HAC formation (40–50 kb). Because the entire procedure of repeat



amplification can be accomplished during ~2–3 weeks and several different arrays can be constructed simultaneously, the new method is readily applicable for the study of the mechanisms of HAC formation.

As proof of principle, we examined the capacity of a 5mer-based 120 kb array generated from a part of the native 11mer HOR to form a HAC. The assembly of *de novo* centromeres from the artificially constructed 5mer-based synthetic array occurred with an efficiency similar to that for native alphoid DNA fragments, suggesting that the existence of a HOR structure for type I arrays in human centromeres is a by-product of human-specific evolutionary mechanisms. The rapid evolution of centromere repeats among different species is consistent with this view. A HOR structure has not yet been detected at the centromeres for most of the organisms for which centromeric tandem repeats have been identified previously (12–14).

Alphoid repeats from different centromeres are not equivalent in their ability to assemble *de novo* centromeres (25,27,38,39), though the sequence differences causing this phenomenon are not yet known. The presence of the CENP-B box is necessary to trigger efficient assembly; however, it is clear that other sequence signals also play a role. These sequences may be unknown motifs that bind centromere proteins or non-specific sequence signal based on epigenetic chromatin assembly. The interplay between such factors and the CENP-B protein may not be equivalent among randomly cloned alphoid repeats. As any nucleotide can be easily changed in a starting repeat sequence before its amplification, the method presented here is a powerful technique for investigations into the sequence requirements for *de novo* centromere/kinetochore seeding. At present, analysis of eight different variants of the 2mer-based alphoid arrays is in progress in our laboratory to reveal the ‘magic’ combination of nucleotides seeding a functional centromere.

There are many other varieties of tandem repeats populating the genomes of eukaryotes. Some of these tandem repeats are known to play important roles in cell function by forming or maintaining specialized chromatin required for chromosome segregation, by stabilizing chromosome ends or by regulating genes. These qualities suggest that tandem repeats may be an important substrate for rapid evolution. As many types of DNA repeats may be similarly amplified, the method described in this paper has more general application to elucidate the role of tandem repeats in the genome. For example, a set of non-alphoid DNA arrays including those generated in this work (i.e. human gamma-8 satellite, mouse major, mouse minor satellite and *Alu*) can be used to address the question of how the composition and length of a tandem repeat array affects heterochromatin formation by targeting the arrays to a structurally defined ectopic chromosomal site by Cre-lox site-specific recombination. Such experiments exploiting the recombination-mediated cassette exchange system (40) are now in progress in our laboratory to contribute to an explanation of the phenomenon of repeat-induced gene silencing (41).

## SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

## ACKNOWLEDGEMENTS

This research was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. We thank K. Yoda (Nagoya University) and N. Nozaki (Kanagawa Dental College) for producing anti-CENP-A antibody, and T. Yen (Fox-Chase Cancer Institute, USA) for the gift of the anti-CENP-E antibody. Funding to pay the Open Access publication charges for this article was provided by National Cancer Institute.

*Conflict of interest statement.* None declared.

## REFERENCES

- Lee, C., Wevrick, R., Fisher, R.B., Ferguson-Smith, M.A. and Lin, C.C. (1997) Human centromeric DNAs. *Hum. Genet.*, **100**, 291–304.
- de Lange, T. (2004) T-loops and the origin of telomeres. *Nature Rev. Mol. Cell. Biol.*, **5**, 323–329.
- Lippman, Z., Gendrel, A.V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B., Kasschau, K.D. et al. (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature*, **430**, 471–476.
- Jasinska, A. and Krzyzosiak, W.J. (2004) Repetitive sequences that shape the human transcriptome. *FEBS Lett.*, **567**, 136–141.
- Li, Y.C., Korol, A.B., Fahima, T. and Nevo, E. (2004) Microsatellites within genes: structure, function, and evolution. *Mol. Biol. Evol.*, **21**, 991–1007.
- Riley, D.E. and Krieger, J.N. (2005) Short tandem repeat (STR) replacements in UTRs and introns suggest an important role for certain STRs in gene expression and disease. *Gene*, **344**, 203–211.
- Mandola, M.V., Stoehlmacher, J., Muller-Weeks, S., Cesarone, G., Yu, M.C., Lenz, H.J. and Ladner, R.D. (2003) A novel single nucleotide polymorphism within the 5′ tandem repeat polymorphism of the thymidylate synthase gene abolishes USF-1 binding and alters transcriptional activity. *Cancer Res.*, **63**, 2898–2904.
- Watanabe, M., Ogawa, Y., Ito, K., Higashihara, M., Kadin, M.E., Abraham, L.J., Watanabe, T. and Horie, R. (2003) AP-1 mediated relief of repressive activity of the CD30 promoter microsatellite in Hodgkin and Reed-Sternberg cells. *Am. J. Pathol.*, **163**, 633–641.
- Everett, C.M. and Wood, N.W. (2004) Trinucleotide repeats and neurodegenerative disease. *Brain*, **127**, 2385–2405.
- Fondon, J.W., III and Garner, H.R. (2004) Molecular origins of rapid and continuous morphological evolution. *Proc. Natl Acad. Sci. USA*, **101**, 18058–18063.
- Sinha, S. and Siggia, E.D. (2005) Sequence turnover and tandem repeats in *cis*-regulatory modules in *Drosophila*. *Mol. Biol. Evol.*, **22**, 874–885.
- Guenatri, M., Bailly, D., Maison, C. and Almouzni, G. (2004) Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. *J. Cell Biol.*, **166**, 493–505.
- Jiang, J., Birchler, J.A., Parrott, W.A. and Dawe, R.K. (2003) A molecular view of plant centromeres. *Trends Plant. Sci.*, **8**, 570–575.
- Sun, X., Le, H.D., Wahlstrom, J.M. and Karpen, G.H. (2003) Sequence analysis of a functional *Drosophila* centromere. *Genome Res.*, **13**, 182–194.
- Ando, S., Yang, H., Nozaki, N., Okazaki, T. and Yoda, K. (2002) CENP-A, -B, and -C chromatin complex that contains the I-type alpha-satellite array constitutes the prekinetochore in HeLa cells. *Mol. Cell. Biol.*, **22**, 2229–2241.
- Spence, J.M., Critcher, R., Ebersole, T.A., Valdivia, M.M., Earnshaw, W.C., Fukagawa, T. and Farr, C.J. (2002) Co-localization of centromere activity, proteins and topoisomerase II within a subdomain of the major human X alpha-satellite array. *EMBO J.*, **21**, 5269–5280.
- Harrington, J.J., Van Bokkelen, G., Mays, R.W., Gustashaw, K. and Willard, H.F. (1997) Formation of *de novo* centromeres and construction of first-generation human artificial microchromosomes. *Nature Genet.*, **15**, 345–355.
- Ikeno, M., Grimes, B., Okazaki, T., Nakano, M., Saitoh, K., Hoshino, H., McGill, N.I., Cooke, H. and Masumoto, H. (1998) Construction of YAC-based mammalian artificial chromosomes. *Nat. Biotechnol.*, **16**, 431–439.

19. Ebersole, T.A., Ross, A., Clark, E., McGill, N., Schindelbauer, D., Cooke, H. and Grimes, B. (2000) Mammalian artificial chromosome formation from circular alphoid input DNA does not require telomere repeats. *Hum. Mol. Genet.*, **9**, 1623–1631.
20. Larin, Z. and Mejia, J.E. (2002) Advances in human artificial chromosome technology. *Trends Genet.*, **18**, 313–319.
21. Kotzamanis, G., Cheung, W., Abdulrazzak, H., Perez-Luz, S., Howe, S., Cooke, H. and Huxley, C. (2005) Construction of human artificial chromosome vectors by recombineering. *Gene*, **351**, 29–38.
22. Ikeno, M., Inagaki, H., Nagata, K., Morita, M., Ichinose, H. and Okazaki, T. (2002) Generation of human artificial chromosomes expressing naturally controlled guanosine triphosphate cyclohydrolase I gene. *Genes Cells*, **7**, 1021–1032.
23. Laner, A., Schwarz, T., Christan, S. and Schindelbauer, D. (2004) Suitability of a CMV/EGFP cassette to monitor stable expression from human artificial chromosomes but not transient transfer in the cells forming viable clones. *Cytogenet. Genome Res.*, **107**, 9–13.
24. Ohzeki, J., Nakano, M., Okada, T. and Masumoto, H. (2002) CENP-B box is required for *de novo* centromere chromatin assembly on human alphoid DNA. *J. Cell Biol.*, **159**, 765–775.
25. Kouprina, N., Ebersole, T., Koriabine, M., Pak, E., Rogozin, I.B., Katoh, M., Oshimura, M., Ogi, K., Peredelchuk, M., Solomon, G. *et al.* (2003) Cloning of human centromeres by transformation-associated recombination in yeast and generation of functional human artificial chromosomes. *Nucleic Acids Res.*, **31**, 922–934.
26. Basu, J., Stromberg, G., Compitello, G., Willard, H.F. and Van Bokkelen, G. (2005) Rapid creation of BAC-based human artificial chromosome vectors by transposition with synthetic alpha-satellite arrays. *Nucleic Acids Res.*, **33**, 587–596.
27. Schueler, M.G., Higgins, A.W., Rudd, M.K., Gustashaw, K. and Willard, H. (2001) Genomic and genetic definition of a functional human centromere. *Science*, **294**, 109–115.
28. Muro, Y., Masumoto, H., Yoda, K., Nozaki, N., Ohashi, M. and Okazaki, T. (1992) Centromere protein B assembles human centromeric alpha-satellite DNA at the 17-bp sequence, CENP-B box. *J. Cell Biol.*, **116**, 585–596.
29. Masumoto, H., Masukata, H., Muro, Y., Nozaki, N. and Okazaki, T. (1989) A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *J. Cell Biol.*, **109**, 1963–1973.
30. Masumoto, H., Yoda, K., Ikeno, M., Kitagawa, K., Muro, Y. and Okazaki, T. (1993) Properties of CENP-B and its target sequence in a satellite DNA. In Vig, B.K. (ed.), *Chromosome and Aneuploidy*. Springer-Verlag, Berlin, pp. 31–43.
31. Leem, S.-H., Noskov, V.N., Park, J.E., Kim, S.I., Larionov, V. and Kouprina, N. (2003) Optimum conditions for selective isolation of genes from complex genomes by transformation-associated recombination cloning. *Nucleic Acids Res.*, **31**, e29.
32. Masumoto, H., Sugimoto, K. and Okazaki, T. (1989) Alphoid satellite DNA is tightly associated with centromere antigens in human chromosomes throughout the cell cycle. *Exp. Cell Res.*, **181**, 181–196.
33. Ikeno, M., Masumoto, H. and Okazaki, T. (1994) Distribution of CENP-B boxes reflected in CREST centromere antigenic sites on long-range alpha-satellite DNA arrays of human chromosome 21. *Hum. Mol. Genet.*, **3**, 1245–1257.
34. Masumoto, H., Ikeno, M., Nakano, M., Okazaki, T., Grimes, B., Cooke, H. and Suzuki, N. (1998) Assay of centromere function using a human artificial chromosome. *Chromosoma*, **107**, 406–416.
35. Yen, T.J., Compton, D.A., Wise, D., Zinkowski, R.P., Brinkley, B.R., Earnshaw, W.C. and Cleveland, D.W. (1991) CENP-E, a novel human centromere-associated protein required for progression from metaphase to anaphase. *EMBO J.*, **10**, 1245–1254.
36. Dean, F.B., Nelson, J.R., Giesler, T.L. and Lasken, R.S. (2001) Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res.*, **11**, 1095–1099.
37. Mizuta, R., Mizuta, M. and Kitamura, D. (2003) Atomic force microscopy analysis of rolling circle amplification of plasmid DNA. *Arch. Histol. Cytol.*, **66**, 175–181.
38. Grimes, B.R., Rhoades, A.A. and Willard, H.F. (2002) Alpha-satellite DNA and vector composition influence rates of human artificial chromosome formation. *Mol. Ther.*, **5**, 798–805.
39. Mejia, J.E., Alazami, A., Willmott, A., Marschall, P., Levy, E., Earnshaw, W.C. and Larin, Z. (2002) Efficiency of *de novo* centromere formation in human artificial chromosomes. *Genomics*, **79**, 297–304.
40. Feng, Y.Q., Seibler, J., Alami, R., Eisen, A., Westerman, K.A., Leboulch, P., Fiering, S. and Bouhassira, E.E. (1999) Site-specific chromosomal integration in mammalian cells: highly efficient CRE recombinase-mediated cassette exchange. *J. Mol. Biol.*, **292**, 779–785.
41. Feng, Y.Q., Lorincz, M.C., Fiering, S., Greally, J.M. and Bouhassira, E.E. (2001) Position effects are influenced by the orientation of a transgene with respect to flanking chromatin. *Mol. Cell Biol.*, **21**, 298–309.