RESEARCH ARTICLE

A satellite-like sequence, representing a "clone gap" in the human genome, was likely involved in the seeding of a novel centromere in macaque

**Lucia Carbone · Pietro D'addabbo ·
Maria Francesca Cardone · Maria Grazia Teti ·
Doriana Misceo · Gery M. Vessere · Pieter J. de Jong ·
Mariano Rocchi**

Abstract Although the human genome sequence is generally considered "finished", the latest assembly (NCBI Build 36.1) still presents a number of gaps. Some of them are defined as "clone gaps" because they separate neighboring contigs. Evolutionary new centromeres are centromeres that repositioned along the chromosome, without marker order variation, during evolution. We have found that one human "clone gap" at 18q21.2 corresponds to an evolutionary new centromere in Old World Monkeys (OWM). The partially sequenced gap revealed a satellite-like structure. DNA stretches of the same satellite were found in the macaque (flanking the chromosome 18 centromere) and in the marmoset (New World Monkey), which was used as an outgroup. These findings strongly suggested that the repeat was present at the time of novel centromere seeding in OWM ancestor. We have provided, therefore, the first instance of a specific sequence hypothesized to have played a role in triggering the emergence of an evolutionary new centromere.

Communicated by: H. Masumoto

L. Carbone · G. M. Vessere · P. J. de Jong
Children's Hospital Research Institute,
Oakland, CA 94609, USA

P. D'addabbo · M. F. Cardone · M. G. Teti · D. Misceo ·
M. Rocchi (✉)
Department of Genetics and Microbiology, University of Bari,
Via Amendola 165/A,
70126 Bari, Italy
e-mail: rocchi@biologia.uniba.it

Introduction

The "finished" human sequence still contains a number of gaps that can be distinguished in "sequence gaps" when spanned by one or more sequencing clones, and "clone gaps" if they separate neighboring contigs (Eichler et al. 2004; Kouprina et al. 2003; Leem et al. 2004).

Clone gaps probably represents unstable sequences that were refractory to transversion by any cloning vector, including fosmid clones that were recently exploited to close some of the gaps (Bovee et al. 2008).

Marker order comparisons between species have shown that the centromeres of orthologous chromosomes can sometimes be embedded in a different genomic context without incurring marker order variation. These repositioned centromeres are indicated as evolutionary new centromeres (ENC). Following seeding, an ENC rapidly acquires the complex organization typical of eukaryotic centromeres (Cardone et al. 2006; Ventura et al. 2007), while the inactivated centromere degrades (Eder et al. 2003; Ventura et al. 2003). The phenomenon appears widespread in the animal kingdom and in plants (for a review, see Marshall et al. 2008).

Within the framework of evolutionary studies aimed at identifying ENCs in primates, we found that the centromere of macaque chromosome 18 is an ENC that emerged in the Old World Monkey (OWM) ancestor. Therefore, the arrangement of the region in humans very likely represented the ancestral organization of the region before ENC seeding. Surprisingly, the mapping of this ENC in humans perfectly corresponded to a clone gap. We hypothesized that the sequences of the gap could have played a role in centromere seeding. The actual length of the gap was

determined and a partial sequencing was achieved. The sequence exhibited a satellite-like structure, with monomers of 111 bp conserved in the great apes and in the marmoset. We also demonstrated that this satellite structure is responsible for the inclonability of the region.

Materials and methods

Fluorescence in situ hybridization (FISH) studies Metaphase preparations were obtained from lymphoblastoid or fibroblast cell lines of the following species. Great apes: common chimpanzee (*Pan troglodytes*, PTR), gorilla (*Gorilla gorilla*, GGO), Borneo orangutan (*Pongo pygmaeus pygmaeus*, PPY); OWM: rhesus monkey (*Macaca mulatta*, MMU, Cercopithecinae), African green monkey (*Cercopithecus aethiops*, CAE, Cercopithecinae), silvered leaf-monkey (*Trachypithecus cristatus*, TCR, Colobinae); New World Monkeys (NWM): common marmoset (*Callithrix jacchus*, CJA, Callitricinae), dusky titi (*Callicebus moloch*, CMO, Callicebinae), squirrel monkey (*Saimiri boliviensis boliviensis*, SBO, Callicebinae).

DNA extraction from BACs was as reported previously (Ventura et al. 2001). FISH experiments were performed essentially as described by Lichter et al. (1990). Digital images were obtained using a Leica DMRXA2 epifluorescence microscope equipped with a cooled CCD camera (Princeton Instruments, Princeton, NJ, USA). Cy3-dCTP, FluorX-dCTP, DEAC, Cy5-dCTP, and DAPI fluorescence signals, detected with specific filters, were recorded separately as gray scale images. Pseudocoloring and merging of images were performed using Adobe Photoshop™ software.

Long-range PCR and sub-cloning PCR products were obtained using primers L1 and R1 reported in Table 2a. LA Takara amplification kit (Takara cat. TAK RR002M) was mainly used for amplification experiments. Standard cloning techniques were used to sub-clones PCR product into pUC19 vector. Briefly, each amplification product was purified using NucleoSpin Extract II kit from Clontech (cat. 636972). Subsequently, ends were repaired using T4 polymerase reaction and ligation reaction was set up. Ligated products were electroporated into DH10B cells (Invitrogen) and plated on ampicillin plates. White/blue selection was used to discriminate the self-ligated vector from successful cloning events.

Zoo blot Standard Southern blot techniques were used to transfer the PCR products from the gel to the membrane. Hybridization was carried out using as a probe an overgo probe (McPherson et al. 2001) designed in the 111U repeat (sequence=TTGAAGTGCCTGCAGTTGTTGTGGCAG ACTGTTCAGTTGC). The experiment was performed at a low stringency conditions: 50% formamide, 6× SSC, 5× Denhardt solution, 0.5 SDS, and 100 μg/ml sonicated herring sperm DNA at 42°C overnight. Washing were done twice at 65°C for 15 min in 2× SSC, 0.1% SDS.

Sequence analysis The dot-plot analysis was performed using GENtle software (http://gentle.magnusmanske.de/; window size=19; mismatch limits=7). The 111SAT consensus analysis was performed using the Consensus software available at http://mobyle.pasteur.fr/cgi-bin/MobylePortal/portal.py?form=consensus.

Results

Chromosome 18 evolutionary history The evolutionary history of human chromosome 18 was investigated by FISH, using an appropriate panel of human BAC clones (Table 1). Analysis of chromosome organization in great apes (chimpanzee, gorilla, and orangutan), in representatives of OWMs (rhesus macaque and vervet monkey, Cercopithecinae; silvered leaf-monkey, Colobinae), and in NWMs (common marmoset, Callitrichinae; dusky titi and squirrel monkey, Callicebinae) showed that, with the exception of humans, the marker order of this chromosome was perfectly conserved if the position of the centromere was not taken into account (Fig. 1a; FISH examples in Fig. 1b). The present results in *Macaca mulatta* perfectly match the rheMac2 (January 2006) sequence assembly now available for this species (Gibbs et al. 2007). The human form is due to a well-known human-specific pericentric inversion of HSA18 short arm (Dennehey et al. 2004; Goidts et al. 2005). At variance with the high marker order conservation, in all the three studied OWM species, the centromere was located in a region corresponding to human 18q21.2, between markers D and E, clearly representing an example of an ENC. The three studied species are representative of both Cercopithecinae and Colobinae subfamilies. Very likely, therefore, the ENC seeding occurred in the OWM ancestor. Using reiterative FISH experiments, the novel centromere was precisely mapped between human BAC clones RP11-61D1 and RP11-289E15 (D2 and E2, respectively, in Fig. 1a; FISH examples in Fig. 1b; sequence position in Table 1). Surprisingly, in humans, the D2 and E2 clones exactly flanked a gap at chr18:50,313,135–50,360,134 (UCSC genome browser, hg18 release).

Evolutionary conservation of the 18q21.2 human sequence In order to validate the human 18q21.2 sequence as a reference sequence with respect to the ancestral organization

Table 1 Panel of BAC clones used to track the evolution of chromosome 18

Code	BAC	Acc. N.	Mapping	UCSC March 2006
A	RP11-78H1	BES	18p11.32	chr18:2,136,811–2,307,213
B	RP11-96I11	BES	18p11.21	chr18:12,904,782–12,904,961
Human centromere				
C	RP11-10G8	BES	18q11.2	chr18:17,274,438–17,431,001
D	RP11-104N11	BES	18q12.2	chr18:33,436,610–33,608,704
D2	RP11-61D1	AC090897	18q21.2	chr18:50,155,761–50,313,129
Gap			18q21.2	chr18:50,313,129–50,360,135
E2	RP11-289E15	AC091135	18q21.2	chr18:50,360,135–50,526,341
E	RP11-153B11	BES	18q21.2	chr18:52,818,203–52,977,905
F	RP11-53N15	BES	18q22.3	chr18:70,195,436–70,195,693
G	RP11-87C15	BES	18q23	chr18:75,965,206–75,965,502
Telomere			18q23	chr18:76,117,153

Letters (code) in the first column corresponds to letters used in Fig. 1. Non-sequenced BACs are placed on the human sequence according to their BAC end sequences (BES) position

of the region at the time of ENC seeding in OWMs, a DNA segment of 2 Mb, centered on the gap, was investigated for conservation against orangutan, mouse, cow, and dog genomes, by visually inspecting the UCSC Comparative Genomics Net tracks (http://genome.ucsc.edu, hg18 assembly). The analysis did not show noteworthy differences.

Sequencing of the 18q21.2 gap The hg18 sequence assembly (UCSC March 2006) reports that the gap is 47 kb in size. The estimation, however, was only based on human–mouse synteny comparison (Nusbaum et al. 2005). In order to define the actual length of the gap, we performed a long-range PCR (LR-PCR) experiment, using primers designed on sequences closely flanking the gap (L1 and R1 in Table 2a). The amplification yielded a fragment of about 9 kb (Fig. 2). The sequencing of the gap was attempted by primer walking, using the LR-PCR-amplified product as template. The sequencing reaction using primer R1 (Table 2a and Fig. 3, which summarizes the sequencing strategy) produced a sequence of about 1 kb. This sequence was aligned against the human genome. It identified the fosmid clone XXfos-87120G11 (38.5 kb, acc. n. AC138661.1) which almost completely overlaps with BAC RP11-289E15 (E2, acc. n. AC091135), with the exception of about 4 kb that protrude into the gap. The second draft (.2) of the AC138661 fosmid sequence contains only its centromeric end (4.262 bp): the first 411 bp overlap with BAC RP11-289E15, while the remaining 3.851 bp protrude inside the gap. This AC138661.2 represents the only chr18_random:1-4,262 (unassigned fragment) of chromosome 18. The sequencing of the telomeric boundary of the gap was resumed using a primer located at the centromeric end of the fosmid sequence (R2 in Table 2a and Fig. 3). All efforts to obtain reads of acceptable quality, however, failed. Sequence analysis revealed that the protruding fosmid sequence was composed of a tandemly repeated DNA (see below). Primer-walking sequencing starting from the centromeric boundary of the gap was more successful. Five rounds of primer walking yielded, in total, about 3.8 kb of sequence (accession number EU074794; primers are reported in Table 2a and Fig. 3). Then the sequence quality dropped dramatically because of the difficulties in designing unique primers, leaving about 1.3 kb of the gap unfinished, as illustrated in Fig. 3a.

Gap sequence organization The dot-plot analysis of AC138661.2 sequence (XXfos-87120G11 fosmid), performed using GENtle software, disclosed that the terminal 2,150 bp of this sequence had a satellite-like structure, with a tandemly repeated monomer of 111 bp (we named 111SAT), in turn composed of three degenerated 37 bp subunits (Supplementary Fig. 1). A consensus sequence of 14 111SAT core monomers present in the AC138661.2 sequence is reported in Supplementary Table 1. The 3.8-kb sequence we obtained by primer-walking experiments (EU074794) was composed of 2.1 kb of unique sequence (centromeric side) and 1.7 kb featuring the same 111SAT structure found in the AC138661.2 sequence. The homology among monomers belonging to EU074794 or AC138661.2 sequences ranged, approximately, from 60% to 85%. We assumed that the unsequenced portion of the gap (1.3 kb) was similarly organized. Therefore, the total length of the satellite DNA stretch was approximately 5 kb in size.

AT content and gene density have been suggested to play a role in neocentromere seeding in human clinical cases and in ENC progression, respectively (Marshall et al. 2008; Ventura et al. 2007). The portions of DNA sequence, composed of the 111-bp repeat in XXfos-87120G11 and in
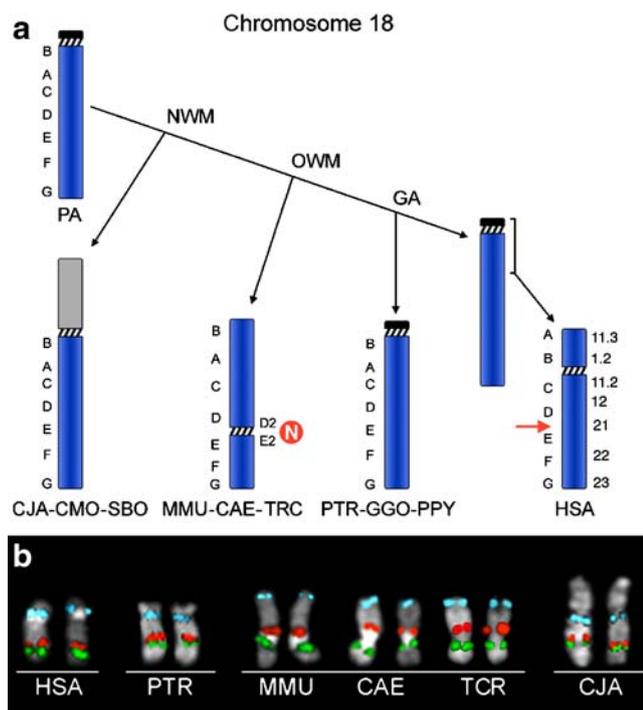
Fig. 1 Evolutionary history of human chromosomes 18. The figure summarizes the evolutionary studies on human chromosomes 18 (**a**), supporting the emergence of ENCs, indicated by the *N* in a *red circle*. The *letters* on the left of each chromosome indicate the BACs used in the study, reported in Table 1. The acronyms indicate: great apes: *PTR* common chimpanzee (*Pan troglodytes*), *GGO* gorilla (*Gorilla gorilla*), *PPY* Borneo orangutan (*Pongo pygmaeus*); OWM: *MMU* rhesus monkey (*Macaca mulatta*), *CAE* vervet monkey (*Cercopithecus aethiops*), *TCR* silvered leaf-monkey (*Trachypithecus cristatus*); NWM: *CJA* common marmoset (*Callithrix jacchus*), *CMO* dusky titi (*Callicebus moloch*), *SBO* squirrel monkey (*Saimiri boliviensis boliviensis*). *PA* stands for primate ancestor. OWM chromosome 18 is upside down to facilitate comparison. *GA* stands for great apes ancestor. **b** Examples of FISH experiments using BAC clones RP11-61D1 (D2, *red*), RP11-289E15 (E2, *green*), and RP11-151D11. D2 and E2 flank the gap in humans and the centromere in OWMs. The RP11-151D11 BAC clone maps, in humans, at 18p11.21, very close to the centromere on 18p, while it gives signals at the telomere in OWM because of the human-specific inversion (see text). Note the large block of alpha satellite repeat revealed by the DAPI staining. Also in (**b**) the OWM chromosome 18 is upside down to facilitate comparison

EU074794, were pooled for AT percentage calculation. They yielded a value of 57%. The AT content was then calculated also for the 50-kb intervals flanking the gap in each side (gap not included). The AT content was 62.8%. C18orf54 (chr18:50,140,895–50,162,381) and C18orf26 (chr18:50,409,388–50,417,722) were identified as the most centromeric and telomeric genes with respect to the gap, respectively. The analysis was performed by querying the human RefSeq-related tracks of UCSC genome browser. The interval between the two genes was 247 kb in size.

Non-clonability of the 18q21.2 gap repeat It is well known that certain DNA repeats are toxic for *Escherichia coli* and

it is therefore not possible to clone them in the commonly used vectors. In order to prove that the cloning difficulties of the gap region were imputable to the 111-bp repeat, we set up the sub-cloning experiments illustrated in Fig. 3b (Table 2b). The region was sub-divided into three amplicons: segment 1 contained unique sequences, segment 2 contained approximately two 111SAT copies, and segment 3 was entirely composed of the 111SAT. Segments 1 and 2 were successfully cloned. Cloning of segment 3 failed despite several attempts. In the attempt to overcome the limitations of classical cloning vectors, we use a different system (Big Easy Cloning kit, Lucigen) based on a novel vector derived from the linear phage N15 of *E. coli*. Because of its unique replication system, this vector (pJAZZ-KA) behaves like a linear plasmid, thus avoiding insert instability due to supercoiling. We performed several cloning attempts in combination with the BigEasy-pTel electrocompetent cells. No colonies were recovered. These results further supported the hypothesis that the absence of clones covering the gap was due to the non-clonability of this stretch of satellite DNA.

Gap sequence evolutionary history

BLAST alignment of the EU074794 sequence against the macaque genome detected a stretch of about 3.1 kb (chr18:47,766,135–47,769,239), of which 2.1 kb were unique sequence and 880 bp were composed of a 111-bp repeat homologous to the human 111SAT (average homology of macaque versus human monomers=76%). The macaque 111SAT was bound, as in humans, to a gap of 47,647 bp (chr18:47,769,245–47,816,891). The human AC138661.2 sequence was aligned, using dot-plot, to the identified macaque sequence. Results clearly showed a similar organization of the human and macaque 111SAT (see Supplementary Fig. 3). The 47-kb gap pinpointed by the AC138661.2 alignment (see above) corresponded to the MMU18 centromere which the rheMac2 assembly missed annotating. Apparently, the rheMac2 assembly did not take into account that several macaque centromeres do not correspond to human centromeres because they are evolutionarily new (Ventura et al. 2007; Roberto et al. 2008). In fact, the macaque sequence facing the gap on the opposite side corresponds to the centromeric BAC end sequence (BES) of the E2 BAC (RP11-289E15).

BLAST analysis using the 111SAT (the stretch present in the EU074794 sequence) picked up in marmoset genome (CJA) the portion 1–725 bp of the unfinished Contig8720 (70.8 kb; calJac1 release, June 2007; note that the marmoset contigs are not assembled in chromosomes yet). Dot-plot analysis, using GENtle, of the ungapped portion of the contig8720:1–4,035 showed that the 1–725 bp stretch had the same satellite-like structure found in humans (Supplementary Fig. 4).

Table 2 Oligo primers used in PCR experiments and in sequencing reactions

Primer name	Symbol	Primer sequence	Sequence acc. n.	Product size (bp)
a. Primers used in sequencing experiments				
61D1_L1	L1	AGCGACCCTTGAACACTCAG chr18:50,313,095–50,313,114	EF458047	
61D1_L2	L2	CTTTAGAGCTGGCTAAATGG	EF458048	
61D1_L3	L3	TATCAACACAAAATGGCAAA	EF458049	
61D1_L4	L4	CTACCACAAAATCCACTTCAC	EF458050	
61D1_L5	L5	TTCTATTGAGACTGCAACCAC	EF458051	
289E15_R	R1	TAACCCCACGTTTTTGGAGT chr18:50,360,168–50,360,187	EF458052	
fosmid_R2	R2	ACTCTTCACTTGCAGTGCTT AC138661.2:3,899–3,918	–	
b. Sub-cloning experiments				
Sub-clone1_R	Sc1_R	CTGATCTAATCAACCCTCATCT		1,394
Sub-clone1_L (61D1_L1)	Sc1_L (L-1)	AGCGACCCTTGAACACTCAG		
Sub-clone2_R	Sc2_R	ATTCAGTAGACGTGGTAGTGGT		1,381
Sub-clone2_L	Sc2_L	CTGAAAAGGAATCTCAAAAGAC		
Sub-clone3_R	Sc3_R	AGAGGTAGTAGTGGTCGGTGTA		1,254
Sub-clone3_L	Sc3_L	CTGAGTCTACAATCACAACAGC		
Sub-clone4_R (289E15_R)	Sc4_R (R-1)	TAACCCCACGTTTTTGGAGT		~4,000
Sub-clone4_L (61D1_L4)	Sc4_L (L-4)	CTACCACAAAATCCACTTCAC		
Positive control_R	PC_R	TGCTACTAGTTTCCTTCCATCT		4,315
Positive control_L	PC_L	TTCAAATGTCACAGGAAAAATA		

In order to directly detect these repeat sequences in the primates under study, we performed several FISH experiments, under different stringency conditions, using long-range PCR products obtained with oligo pairs L-1/R-1 and R-1/L-4 (Table 2). However, no specific signal was obtained. The failure can be attributed to the relatively small size of the target sequences.

The evolutionary history of the satellite segment was then investigated by PCR and Southern blot experiments on amplified products. Long-range PCR, using L1 and R1 primers (Table 2a), were performed on DNA from representatives of Prosimians (ring-tailed lemur, *Lemur catta*, LCA), New World monkeys (common marmoset, CJA), Old World monkey (rhesus macaque, MMU), and great apes (orangutan, PPY; gorilla, GGO; chimpanzee, PTR). Bovine (*Bos taurus*, BTA), was used as an outgroup. A 9-kb PCR product, comparable to the human amplification product, was obtained in great apes (Fig. 4). A slightly smaller and weaker band was obtained in marmoset, while an amplification band of about 4 kb was observed in lemur. No amplification product was seen in macaque. The amplified products were digested with the restriction enzyme *Pst*I, which is supposed to cut on both sides of the 111-bp monomer, and blotted on nylon membranes. The hybridization was carried out at low stringency using an overgo probe spanning the whole monomer sequence (see "Material and methods" section). Figure 5 shows the digestion products (a) and the hybridization pattern (b). Clear bands were present in great apes but absent in all remaining analyzed species, including the marmoset.

Discussion

In the present paper, we have reported an intriguing example of a human clone gap, at 18q21.2, that corresponds to an ENC in OWMs. Two BAC clones, precisely delimiting the gap in humans, were found to perfectly flank the ENC of macaque chromosome 18. We attempted the sequencing of the 18q21.2 human gap in order to identify the sequences supposedly responsible for the inclonability of the region and, potentially, for the ENC emergence. Although we were not able to achieve a complete sequence of the gap, the partial sequence strongly indicated that the gap core was composed of approximately 5 kb of tandemly repeated 111-bp unit (111SAT). Sub-cloning experiments indicated that this repeat was responsible for the inclonability of the region.

The BLAST analysis of the human EU074794 repeat against macaque and marmoset genomes, using low stringent criteria, identified sequences that mapped exactly to the region corresponding to the human gap, including a DNA stretch composed of the 111SAT repeat. In macaque, these sequences precisely flanked the evolutionary new centromere. We also tracked in several primate species the evolution of the 111SAT repeat, by Southern blotting analysis, on genomic DNA amplified using oligos designed on the sequences flanking the gap (L1 and R1) (Fig. 4). As expected, the macaque genomic DNA failed to amplify because of the intervening centromere. Southern blotting on amplified products digested with *Pst*I yielded positive bands in humans and great apes. No hybridization bands

Fig. 2 Long-range PCR. Long-range-PCR-amplified product of L1 and R1 oligo primers (see Table 2), designed on sequences flanking the gap at 18q21.2, using human genomic DNA as template (*lane 2*). *Lane 1*: DNA size markers
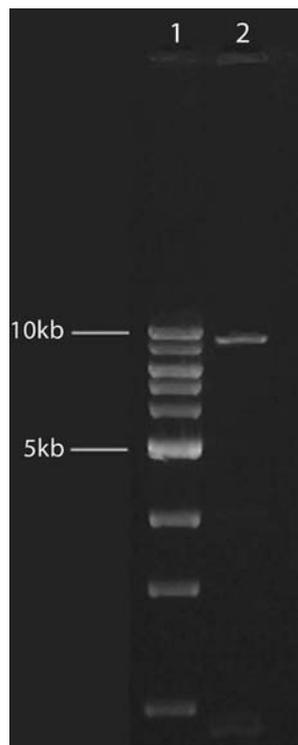
were present in marmoset (CJA), lemur (LCA), and cattle (BTA). The absence of signals was probably due to the inadequacy of the technique to detect sequences with a relatively low similarity.

Our bioinformatics investigation on the macaque and marmoset assemblies established that the 111SAT was present in OWM ancestor at the time of neocentromere seeding. The repeat, according to the present rheMac2 macaque assembly, faces the gap corresponding to the MMU18 centromere. The perfect correspondence of the neocentromere seeding point with the 111SAT mapping

suggests that 111SAT could have been involved in triggering the emergence of the evolutionary new centromere. The emergence of evolutionary and clinical neocentromeres has been hypothesized to be epigenetic in nature, that is, not accompanied by any sequence transposition (Marshall et al. 2008), and epigenetic factors are supposed to play a fundamental role in the functioning of a normal centromere (Gieni et al. 2008).

The repositioned centromere recruited a large block of alpha satellite DNA, which then apparently acquired the centromeric function. A similar process have affected all the nine repositioned centromeres in macaque (Ventura et al. 2007). The recruiting mechanism remains, however, obscure. The seeding of a human neocentromere in the heterochromatic block of chromosome Y has been reported by Tyler-Smith et al. (1999). To our knowledge, there are no prior examples specifically linking a satellite DNA to the emergence of a neocentromere.

In humans, several cases of clinical neocentromeres have been described (Marshall et al. 2008). Seven of them have been precisely mapped using ChIP-on-chip technology (reviewed in Marshall et al. 2008; Capozzi et al. 2008), but a satellite DNA has never been implicated in the neocentromere emergence. The only shared feature was a relatively high AT content (59.9–66.1%), with respect to the average 59% of the entire human genome (Lander et al. 2001) and to the AT content of the alpha satellite DNA (61.4%). The AT content around the gap was relatively high (higher than 62%), while the 111-bp repeat, composing most of the gap, was 57%, which is relatively low. Very likely, it was the satellite DNA that played a major role in centromere seeding.

Domains at 3q26, 13q21, and 9q33.1 have been shown to retain latent, inherent potential to form novel centro-
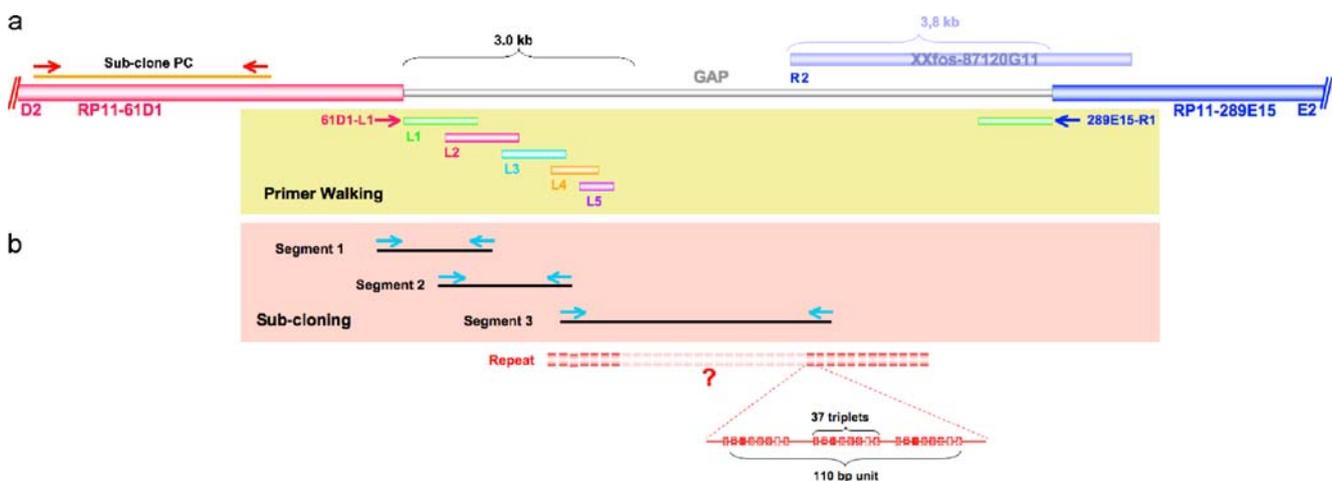
Fig. 3 Map of the region around the 18q21.2 gap. **a** shows the primer-walking strategy starting from the two BAC ends flanking the gap on chromosome 18, along with the distribution of the 111U repeat and its structure. **b** graphically shows the sub-cloning experiments performed to test the clonability of the region. Primers are reported in Table 2. For details see text

Fig. 4 Amplification product in different species, using L1 and R2 primers (Table 2). For acronyms indicating the different species, see legend to Fig. 1. For details see text
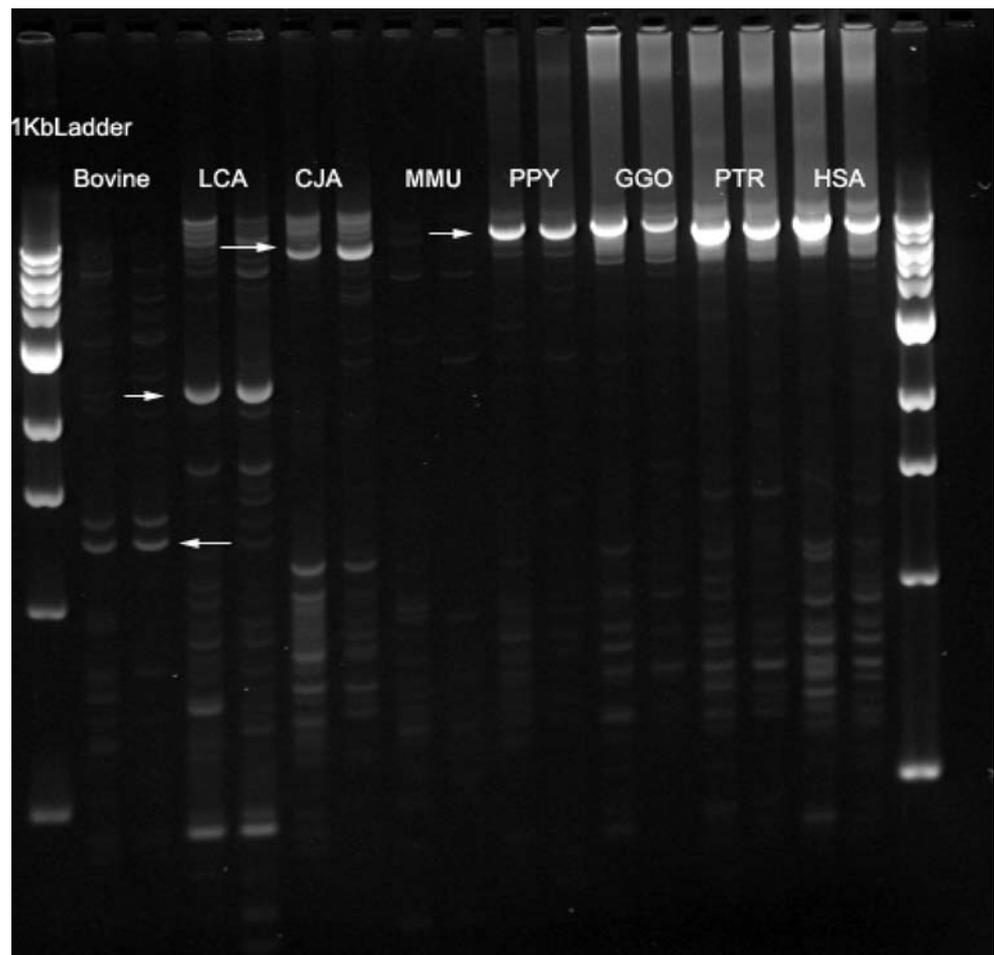
meres, both in evolution and in clinical cases (Capozzi et al. 2008; Cardone et al. 2006; Ventura et al. 2004). Clinical cases showing a neocentromere at 18q21.2 have never been reported, contrary to theoretical expectations suggested by the present findings. It can be argued, however, that most of

the supernumerary fragments bearing a neocentromere are organized as inverted duplications (invdup) leading to a tetrasomy of the region, a condition that, in case of chromosome 18, could be incompatible with a normal development.
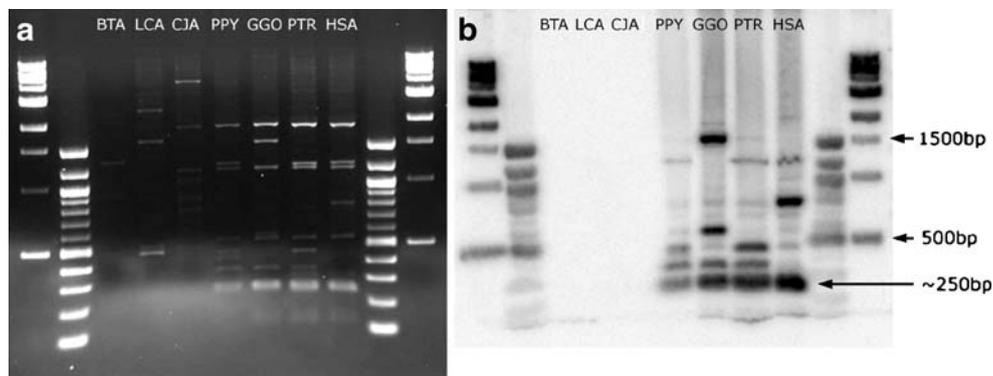
Fig. 5 The amplified products shown in Fig. 4 were digested with *Pst*I restriction enzyme (**a**) and hybridized using, as a probe, the overgo probe designed in the 111U repeat (see "Materials and methods" section) (**b**). Negative bands in HSA, PTR, GGO, and PPY very likely represent DNA stretches external to the repeated region. For acronyms indicating the different species, see legend to Fig. 1

References

Bovee D, Zhou Y, Haugen E, Wu Z, Hayden HS, Gillett W, Tuzun E, Cooper GM, Sampas N, Phelps K, Levy R, Morrison VA, Sprague J, Jewett D, Buckley D, Subramaniam S, Chang J, Smith DR, Olson MV, Eichler EE, Kaul R (2008) Closing gaps in the human genome with fosmid resources generated from multiple individuals. Nat Genet 40:96–101

Capozzi O, Purgato S, Verdun di Cantogno L, Grosso E, Ciccone R, Zuffardi O, Della Valle G, Rocchi M (2008) Evolutionary and clinical neocentromeres: two faces of the same coin. Chromosoma 117:339–344

Cardone MF, Alonso A, Pazienza M, Ventura M, Montemurro G, Carbone L, de Jong PJ, Stanyon R, D'Addabbo P, Archidiacono N, She X, Eichler EE, Warburton PE, Rocchi M (2006) Independent centromere formation in a capricious, gene-free domain of chromosome 13q21 in Old World monkeys and pigs. Genome Biol (www) 7:R91

Dennehey BK, Gutches DG, McConkey EH, Krauter KS (2004) Inversion, duplication, and changes in gene context are associated with human chromosome 18 evolution. Genomics 83:493–501

Eder V, Ventura M, Ianigro M, Teti M, Rocchi M, Archidiacono N (2003) Chromosome 6 phylogeny in primates and centromere repositioning. Mol Biol Evol 20:1506–1512

Eichler EE, Clark RA, She X (2004) An assessment of the sequence gaps: unfinished business in a finished human genome. Nat Rev Genet 5:345–354

Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, Mardis ER, Remington KA, Strausberg RL, Venter JC, Wilson RK, Batzer MA, Bustamante CD, Eichler EE, Hahn MW, Hardison RC, Makova KD, Miller W, Milosavljevic A, Palermo RE, Siepel A, Sikela JM, Attaway T, Bell S, Bernard KE, Buhay CJ, Chandrabose MN, Dao M, Davis C, Delehaunty KD, Ding Y, Dinh HH, Dugan-Rocha S, Fulton LA, Gabisi RA, Garner TT, Godfrey J, Hawes AC, Hernandez J, Hines S, Holder M, Hume J, Jhangiani SN, Joshi V, Khan ZM, Kirkness EF, Cree A, Fowler RG, Lee S, Lewis LR, Li Z, Liu YS, Moore SM, Muzny D, Nazareth LV, Ngo DN, Okwuonu GO, Pai G, Parker D, Paul HA, Pfannkoch C, Pohl CS, Rogers YH, Ruiz SJ, Sabo A, Santibanez J, Schneider BW, Smith SM, Sodergren E, Svatek AF, Utterback TR, Vattathil S, Warren W, White CS, Chinwalla AT, Feng Y, Halpern AL, Hillier LW, Huang X, Minx P, Nelson JO, Pepin KH, Qin X, Sutton GG, Venter E, Walenz BP, Wallis JW, Worley KC, Yang SP, Jones SM, Marra MA, Rocchi M, Schein JE, Baertsch R, Clarke L, Csuros M, Glasscock J, Harris RA, Havlak P, Jackson AR, Jiang H, Liu Y, Messina DN, Shen Y, Song HX, Wylie T, Zhang L, Birney E, Han K, Konkel MK, Lee J, Smit AF, Ullmer B, Wang H, Xing J, Burhans R, Cheng Z, Karro JE, Ma J, Raney B, She X, Cox MJ, Demuth JP, Dumas LJ, Han SG, Hopkins J, Karimpour-Fard A, Kim YH, Pollack JR, Vinar T, Addo-Quaye C, Degenhardt J, Denby A, Hubisz MJ, Indap A, Kosiol C, Lahn BT, Lawson HA, Marklein A, Nielsen R, Vallender EJ, Clark AG, Ferguson B, Hernandez RD, Hirani K, Kehrer-Sawatzki H, Kolb J, Patil S, Pu LL, Ren Y, Smith DG, Wheeler DA, Schenck I, Ball EV, Chen R, Cooper DN, Giardine B, Hsu F, Kent WJ, Lesk A, Nelson DL, O'Brien WE, Prufer K, Stenson PD, Wallace JC, Ke H, Liu XM, Wang P, Xiang AP, Yang F, Barber GP, Haussler D, Karolchik D, Kern AD, Kuhn RM, Smith KE, Zwieg AS (2007) Evolutionary and biomedical insights from the rhesus macaque genome. Science 316:222–234

Gieni RS, Chan GK, Hendzel MJ (2008) Epigenetics regulate centromere formation and kinetochore function. J Cell Biochem 104:2027–2039

Goidts V, Szamalek JM, de Jong PJ, Cooper DN, Chuzhanova N, Hameister H, Kehrer-Sawatzki H (2005) Independent intra-chromosomal recombination events underlie the pericentric inversions of chimpanzee and gorilla chromosomes homologous to human chromosome 16. Genome Res 15:1232–1242

Kouprina N, Leem SH, Solomon G, Ly A, Koriabine M, Otstot J, Pak E, Dutra A, Zhao S, Barrett JC, Larionov V (2003) Segments missing from the draft human genome sequence can be isolated by transformation-associated recombination cloning in yeast. EMBO Rep 4:257–262

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, Szustakowski J, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

Leem SH, Kouprina N, Grimwood J, Kim JH, Mullokandov M, Yoon YH, Chae JY, Morgan J, Lucas S, Richardson P, Detter C, Glavina T, Rubin E, Barrett JC, Larionov V (2004) Closing the gaps on human chromosome 19 revealed genes with a high

density of repetitive tandemly arrayed elements. Genome Res 14:239–246

Lichter P, Tang Chang C-J, Call K, Hermanson G, Evans GA, Housman D, Ward DC (1990) High resolution mapping of human chromosomes 11 by in situ hybridization with cosmid clones. Science 247:64–69

Marshall OJ, Chueh AC, Wong LH, Choo KH (2008) Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. Am J Hum Genet 82:261–282

McPherson JD, Marra M, Hillier L, Waterston RH, Chinwalla A, Wallis J, Sekhon M, Wylie K, Mardis ER, Wilson RK, Fulton R, Kucaba TA, Wagner-McPherson C, Barbazuk WB, Gregory SG, Humphray SJ, French L, Evans RS, Bethel G, Whittaker A, Holden JL, McCann OT, Dunham A, Soderlund C, Scott CE, Bentley DR, Schuler G, Chen HC, c, Jang W, Green ED, Idol JR, Maduro VV, Montgomery KT, Lee E, Miller A, Emerling S, Kuherlapati R, Gibbs R, Scherer S, Gorrell JH, Sodergren E, Clerc-Blankenburg K, Tabor P, Naylor S, Garcia D, de Jong PJ, Catanese JJ, Nowak N, Osoegawa K, Qin S, Rowen L, Madan A, Dors M, Hood L, Trask B, Friedman C, Massa H, Cheung VG, Kirsch IR, Reid T, Yonescu R, Weissenbach J, Bruls T, Heilig R, Branscomb E, Olsen A, Doggett N, Cheng JF, Hawkins T, Myers RM, Shang J, Ramirez L, Schmutz J, Velasquez O, Dixon K, Stone NE, Cox DR, Haussler D, Kent WJ, Furey T, Rogic S, Kennedy S, Jones S, Rosenthal A, Wen G, Schilhabel M, Gloeckner G, Nyakatura G, Siebert R, Schlegelberger B, Korenberg J, Chen XN, Fujiyama A, Hattori M, Toyoda A, Yada T, Park HS, Sakaki Y, Shimizu N, Asakawa S, Kawasaki K, Sasaki T, Shintani A, Shimizu A, Shibuya K, Kudoh J, Minoshima S, Ramser J, Seranski P, Hoff C, Poustka A, Reinhardt R, Lehrach H (2001) A physical map of the human genome. Nature 409:934–941

Nusbaum C, Zody MC, Borowsky ML, Kamal M, Kodira CD, Taylor TD, Whittaker CA, Chang JL, Cuomo CA, Dewar K, FitzGerald MG, Yang X, Abouelleil A, Allen NR, Anderson S, Bloom T, Bugalter B, Butler J, Cook A, DeCaprio D, Engels R, Garber M, Gnirke A, Hafez N, Hall JL, Norman CH, Itoh T, Jaffe DB, Kuroki Y, Lehoczky J, Lui A, Macdonald P, Mauceli E, Mikkelsen TS, Naylor JW, Nicol R, Nguyen C, Noguchi H, O'Leary SB, Piqani B, Smith CL, Talamas JA, Topham K, Totoki Y, Toyoda A, Wain HM, Young SK, Zeng Q, Zimmer AR, Fujiyama A, Hattori M, Birren BW, Sakaki Y, Lander ES (2005) DNA sequence and analysis of human chromosome 18. Nature 437:551–555

Roberto R, Misceo D, D'Addabbo P, Archidiacono N, Rocchi M (2008) Refinement of macaque synteny arrangement with respect to the official rheMac2 macaque sequence assembly. Chromosome Res 16:977–985

Tyler-Smith C, Gimelli G, Giglio S, Floridia G, Pandya A, Terzoli G, Warburton PE, Earnshaw WC, Zuffardi O (1999) Transmission of a fully functional human neocentromere through three generations. Am J Hum Genet 64:1440–1444

Ventura M, Archidiacono N, Rocchi M (2001) Centromere emergence in evolution. Genome Res 11:595–599

Ventura M, Mudge JM, Palumbo V, Burn S, Blennow E, Pierluigi M, Giorda R, Zuffardi O, Archidiacono N, Jackson MS, Rocchi M (2003) Neocentromeres in 15q24-26 map to duplicons which flanked an ancestral centromere in 15q25. Genome Res 13:2059–2068

Ventura M, Weigl S, Carbone L, Cardone MF, Misceo D, Teti M, D'Addabbo P, Wandall A, Björck E, de Jong P, She X, Eichler EE, Archidiacono N, Rocchi M (2004) Recurrent sites for new centromere seeding. Genome Res 14:1696–1703

Ventura M, Antonacci F, Cardone MF, Stanyon R, D'Addabbo P, Cellamare A, Sprague LJ, Eichler EE, Archidiacono N, Rocchi M (2007) Evolutionary formation of new centromeres in macaque. Science 316:243–246