Chromosoma (1996) 105:180–189

CHROMOSOMA

Analysis of the organisation and localisation of the FSHD-associated tandem array in primates: implications for the origin and evolution of the 3.3 kb repeat family

Lorraine N. Clark[1, *], **Udo Koehler**[2], **David C. Ward**[3], **Johannes Wienberg**[2, *], **Jane E. Hewitt**[1]

[1] School of Biological Sciences, 3.239 Stopford Building, The University of Manchester, Oxford Road, Manchester M13 9PT, UK
[2] Institute of Anthropology and Human Genetics, Ludwig-Maximilian University, Munich, Germany
[3] Yale University School of Medicine, Department of Genetics, New Haven, CT 06510, USA

Received: 4 June 1996 / Accepted: 18 June 1996

Abstract. The D4Z4 locus is a polymorphic tandem repeat sequence on human chromosome 4q35. This locus is implicated in the neuromuscular disorder facioscapulohumeral muscular dystrophy (FSHD). The majority of sporadic cases of FSHD are associated with *de novo* DNA deletions within D4Z4. However, it is still not known how this rearrangement causes FSHD. Although the repeat contains homeobox sequences, despite exhaustive searching, no transcript from this locus has been identified. Therefore, it has been proposed that the deletion may invoke a position effect on a nearby gene. In order to try to understand the role of the D4Z4 repeat in this disease, we decided to investigate its conservation in other species. In this study, the long-range organisation and localisation of loci homologous to D4Z4 were investigated in primates using Southern blot analysis, pulsed field gel electrophoresis and fluorescence *in situ* hybridisation. In humans, probes to D4Z4 identify, in addition to the 4q35 locus, a closely related tandem repeat at 10qter and many related repeat loci mapping to the acrocentric chromosomes; a similar pattern was seen in all the great apes. In Old World monkeys, however, only one locus was detected in addition to that on the homologue of human chromosome 4, suggesting that the D4Z4 locus may have originated directly from the progenitor locus. The finding that tandem arrays closely related to D4Z4 have been maintained at loci homologous to human chromosome 4q35–qter in apes and Old World monkeys suggests a functionally important role for these sequences.

* *Present address:* Department of Neurology, School of Medicine, San Francisco General Hospital, San Francisco, USA
** *Present address:* Department of Pathology, Tennis Court Road, Cambridge, UK

Edited by: H.F. Willard

Correspondence to: J. Hewitt (e-mail: jhewitt@hgmp.mrc.ac.uk)

Introduction

Facioscapulohumeral muscular dystrophy (FSHD) is an autosomal dominant, progressive neuromuscular disorder with an estimated prevalence of 1–5 per 100000 (Lunt and Harper 1991). Genetic and physical mapping strategies have identified a polymorphic 3.3 kb tandem repeat (D4Z4) on human chromosome 4q35 that is tightly linked to FSHD (Wijmenga et al. 1992; Passos-Bueno et al. 1993; Upadhyaya et al. 1993; Weiffenbach et al. 1993; Wright et al. 1993; Wijmenga et al. 1994). In unaffected individuals the *Eco*RI fragment containing D4Z4 varies in size between approximately 50 and 320 kb, while in familial cases of FSHD linked to chromosome 4 the disease co-segregates with a fragment below this range in size. The probe p13E-11 (D4F104S1), which maps just proximal to D4Z4 on the same *Eco*RI fragment, detects *de novo* DNA rearrangements in the majority of sporadic FSHD cases (Wijmenga et al. 1992; Passos-Bueno et al. 1993; Upadhyaya et al. 1993; Weifenbach et al. 1993; Wijmenga et al. 1994). It has been demonstrated that the DNA rearrangement involves deletion of an integral number of 3.3 kb repeat units (van Deutekom et al. 1993).

D4Z4 was originally isolated as part of a cosmid (13E) identified by an oligonucleotide probe to helix 3 of the homeodomain (Wijmenga et al. 1992). Each copy of the 3.3 kb repeat contains two homeobox sequences and, in addition, two different classes of GC-rich repetitive DNA (Hewitt et al. 1994; Winokur et al. 1994). hhspm3, a member of a low copy human repeat family previously identified by Zhang et al. (1987), and LSau, a middle repetitive DNA family (Agresti et al. 1989; Meneveri et al. 1993). In humans LSau is closely associated with β satellite DNA and is found in heterochromatic regions; in particular, the short arms of the acrocentric chromosomes, the pericentromeric region of chromosomes 1, 3 and 9, and the proximal euchromatic region (q11.21–pter) of the Y chromosome. This repeat has recently been demonstrated to be conserved in great apes (Meneveri et al. 1995).

Southern blot hybridisation and fluorescence *in situ* hybridisation (FISH) data suggested that there are repeat sequences similar to D4Z4 at predominantly heterochromatic regions in the human genome, such as the short arms of the acrocentric chromosomes (Hewitt et al. 1994; Winokur et al. 1994). Recently we described several yeast artificial chromosome (YAC) and cDNA clones that contain D4Z4-like sequences and map to the acrocentric chromosomes, showing that this cross-hybridisation detected by FISH is due to similar homeobox-containing repeats (Hewitt et al. 1994; Lyle et al. 1995). We have proposed the name 3.3 kb repeat for this family, after the suggestion made by Tyler-Smith and Willard (1993) to simplify the naming of repetitive DNA and after the length of the first member to be described (D4Z4).

D4Z4 is strongly implicated as having a causative role in the aetiology of FSHD: deletions within this locus are associated with the disease. However, the precise disease mechanism has yet to be elucidated. Although an open reading frame has been identified within D4Z4 that could encode a homeodomain-containing protein, there is, to date, no evidence that it is transcribed (Hewitt et al. 1994; Winokur et al. 1994; Lyle et al. 1995). The structure of D4Z4 and the distribution of related loci in the human genome suggest that these repeats are associated with heterochromatin. It has been postulated that deletion of D4Z4 sequences could produce a position effect that alters the activity of the FSHD gene (Hewitt et al. 1994; Winokur et al. 1994; Lyle et al. 1995). Recently, a gene has been identified that maps 100 kb proximal to the D4Z4 repeat (van Deutekom et al. 1996). No changes in FRG1 expression levels could be detected in FSHD patients versus controls. To date, no other genes have been identified in this region of chromosome 4q35. Therefore, a more detailed analysis of the D4Z4 repeat may be fruitful.

The aim of this work was to investigate the evolution of this repeat family in order to gain insights into its role in FSHD. We have investigated the conservation and long-range organisation of the individual sequence elements that comprise D4Z4 in the great apes and rhesus macaque, an Old World monkey. In addition, the 3.3 kb repeat family loci has been localised in primates using FISH.

Materials and methods

Genomic DNA samples. Human and macaque genomic DNAs used in this study were isolated from peripheral blood or tissue samples. Great ape genomic DNA was prepared from lymphoblast cell lines or purchased from BIOS Laboratories (New Haven). Chimpanzee peripheral lymphoblast (EB 176), gorilla peripheral lymphoblast (EB JC) and orang-utan peripheral lymphoblast (EB185 JC) cell lines were obtained from the European Collection of Animal Cell Cultures. Genomic DNA was isolated according to Miller et al. (1988).

Southern blot analysis. Approximately 8 μg of genomic DNA was digested with the appropriate restriction enzyme according to the manufacturer's instructions (Boehringer-Mannheim). Agarose blocks containing high molecular weight genomic DNA were equilibrated for 1 h at 37° C with 0.5 ml of the appropriate restriction enzyme buffer before digestion with 200 U of *Eco*RI, *Hind*III or *Kpn*I for 16 h at 37° C. For separation of fragments <25 kb, standard horizontal gel electrophoresis was used. To resolve DNA fragments >25 kb, electrophoresis was performed at 14° C in 0.5×TBE using a BioRad Chef DRII apparatus. Switch times are given in figure legends. Lambda DNA concatamers (NEB), lambda DNA digested with *Hind*III (Gibco BRL) and DNA of yeast strain YPH80 (BioRad) were run as markers. DNA was transferred to PALL B (PALL Biodyne) by alkaline Southern blotting for 16–48 h.

DNA probes. The isolation of cosmid 13E has been described previously (Wijmenga et al. 1992). The probe 9B6A, which contains the double-homeobox sequence from D4Z4, has been described previously (Wright et al. 1993). The probe K3.3 encompasses the complete D4Z4 repeat unit and was isolated by gel purification at 3.3 kb fragment from a *Kpn*I digest of cosmid 13E. Probes containing other sequence motifs from D4Z4 were produced by subcloning appropriate fragments from cosmid 13E (Fig. 1) into pBluescript SK^+ (Stratagene). B6P1 is a 320 bp *Bam*HI/*Pst*I subclone that contains the LSau sequence. Probe BS1 is a 460 bp *Bam*HI/*Sac*II subclone containing the hhspm3 sequence. Probes were labelled with [α^{32}P]dCTP by the random primer technique. Hybridisation was performed in Church buffer (Church and Gilbert 1984) at 65° C for 16–24 h, except for BS1, which was hybridised at 75° C. Unless stated otherwise, filters were washed at a stringency of 2×SSC, 0.1% SDS at 65° C; followed by autoradiography for 2–5 days at –70° C using medical X-ray film (Fuji) with intensifying screens.

Fluorescent in situ hybridisation. The D4Z4-containing probes used were cosmid 13E (Wijmenga et al. 1992), YAC clone 25C2E (Wright et al. 1993), and cosmid C13, which was isolated from the YAC 25C2E and spans the D4Z4 repeat (T. Wright, unpublished data). For each species, all clones gave a similar distribution pattern of signals when used for FISH. Genomic synteny at the DNA level has been demonstrated between human chromosome 4 and chromosome 3 of pigmy chimpanzee, gorilla and orang-utan by chromosome *in situ* suppression hybridisation (Jauch et al. 1992). Therefore, in order to identify the great ape homologues of human chromosome 4, a paint produced from a human chromosome 4-specific phage library (ATCC number LA04NSO2) was used in combination with the D4Z4 probes. Two micrograms of YAC, cosmid or plasmid DNA was labelled by nick-translation as described previously (Ried et al. 1993). Cosmid and plasmid biotin-labelling reactions were incubated for 2–3 h at 15° C and those of YACs for 4–6 h at 15° C. Two micrograms of the human chromosome 4-specific DNA library was labelled with digoxigenin. Hybridisation and image analysis were performed as described previously (Ried et al. 1993).

Results

All the sequence elements within D4Z4 are conserved and closely associated in the great apes

The D4Z4 probes used in this study are shown in Fig. 1. The restriction enzyme *Kpn*I cuts once within every copy of the D4Z4 repeat, reducing the D4Z4 locus to multiple copies of a 3.3 kb fragment (Wright et al. 1993). Therefore, we initially used *Kpn*I-digested DNA to analyse the evolutionary conservation of D4Z4 in the great apes. The probe K3.3 was hybridised to genomic DNA from human (HSA), chimpanzee (PTR), gorilla (GGO), orang-utan (PPY) and an Old World monkey, baboon (PHA). A strongly hybridising 3.3 kb *Kpn*I frag-

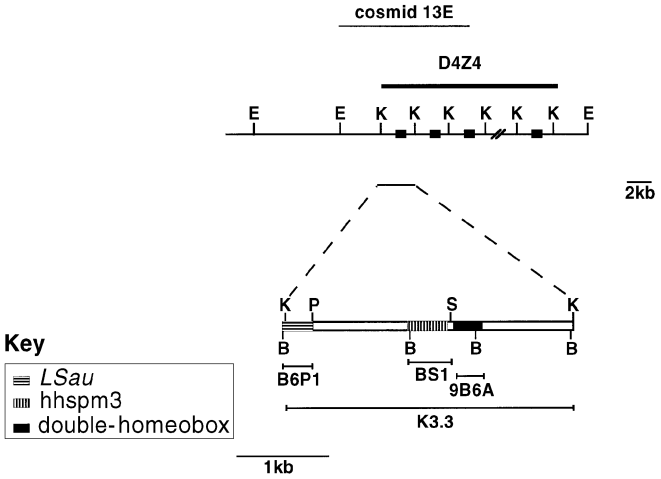

Fig. 1. A schematic of the D4Z4 region showing an expanded map of one complete D4Z4 repeat unit. The location of cosmid 13E is indicated. *Filled boxes* indicate the positions of the double-homeobox sequences within D4Z4. The positions of sequence motifs in D4Z4 are shown; refer to key for identity. *Horizontal bars* indicate the positions of probes used in this study. Restriction enzyme sites: *B Bam*HI, *E Eco*RI, *K Kpn*I, *P Pst*I, *S Sac*II. For clarity, not all *Sac*II and *Pst*I restriction sites are indicated

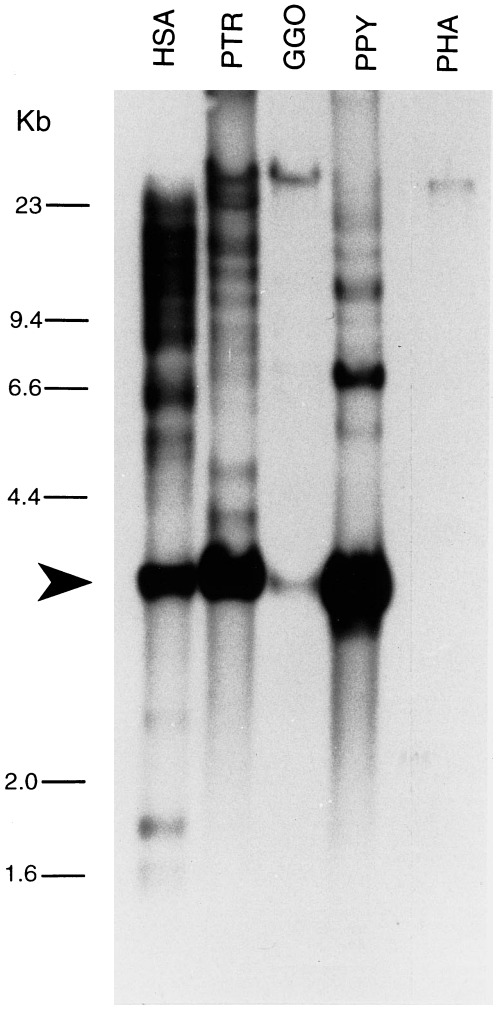

Fig. 2. A Southern blot containing 8 μg *Kpn*I-digested genomic DNA from human (*HSA*), chimpanzee (*PTR*), gorilla (*GGO*), orang-utan (*PPY*), and baboon (*PHA*) was hybridised with the K3.3 probe. The filter was washed in 2×SSC, at 65° C and exposed for 2 days. The *arrowhead* indicates the 3.3 kb fragment corresponding to D4Z4

ment, the same size as in human DNA was detected in all the great apes but not in baboon, which had a cross-hybridising fragment of >23 kb (Fig. 2). In all the great apes additional hybridising bands were seen. The intensity of the 3.3 kb fragments in human, chimpanzee and orang-utan suggests that multiple copies of this fragment are present.

The conservation of the homeobox, hhspm3 and LSau sequence elements of the D4Z4 repeat in the great apes was investigated further using Southern blot analysis of *Eco*RI- and *Pst*I-digested genomic DNA (Figs. 3, 4). The homeobox probe 9B6A detected many *Eco*RI fragments (some at the limit of resolution) in chimpanzee, gorilla and orang-utan. The LSau probe, B6P1, gave a similar pattern of hybridisation to the probe 9B6A, suggesting these two sequence motifs are closely associated. The result in orang-utan was particularly striking: as with 9B6A two strongly hybridising fragments of 6 and 7.5 kb were detected. The hhspm3 probe, BS1, also gave many cross-hybridising fragments. Again, there were fragments in common with the homeobox probe, in particular the 6 kb *Eco*RI fragment in orang-utan. Thus the homeobox, hhspm3 and LSau elements from D4Z4 are conserved in the genomes of the great apes.

To investigate further the association of these D4Z4 sequence motifs the enzyme *Pst*I, which custs within the human D4Z4 repeat unit, was then used (Fig. 4). Many positive fragments were detected using 9B6A. Within D4Z4, the homeobox motif maps to a 1.4 kb *Pst*I fragment; although a band of this size was detected by the 9B6A probe in human DNA, it was not seen in any other species. However, a strongly cross hybridising fragment of approximately 2 kb is conserved in all the great apes. The hhspm3 probe produced a pattern of hybridisation that was almost identical to that seen with 9B6A. The hhspm3 sequence maps to the same 1.4 kb *Pst*I fragment within D4Z4 as the homeobox motif. Again, a band of this size was seen only in human DNA with a strongly hybridising fragment of approximately 2 kb present in all the great apes. Within D4Z4, LSau maps to 634 bp *Pst*I fragment; the probe B6P1 detected in a band of this size in all species, in addition to other fragments.

From these results, it appears that chimpanzee, orang-utan and gorilla all contain a locus or loci that is/are very similar to D4Z4: probes to the repeat identify a 3.3 kb *Kpn*I fragment in all species studied in the sequence elements that comprise D4Z4 are conserved and found in close association. More than one hybridising fragment was detected by each probe, suggesting that the great apes (in common with man) also contain additional members of the 3.3 kb repeat family.

The human D4Z4 locus is contained within a polymorphic *Eco*RI fragment that varies in size from approximately 50 to greater than 300 kb and pulsed field gel

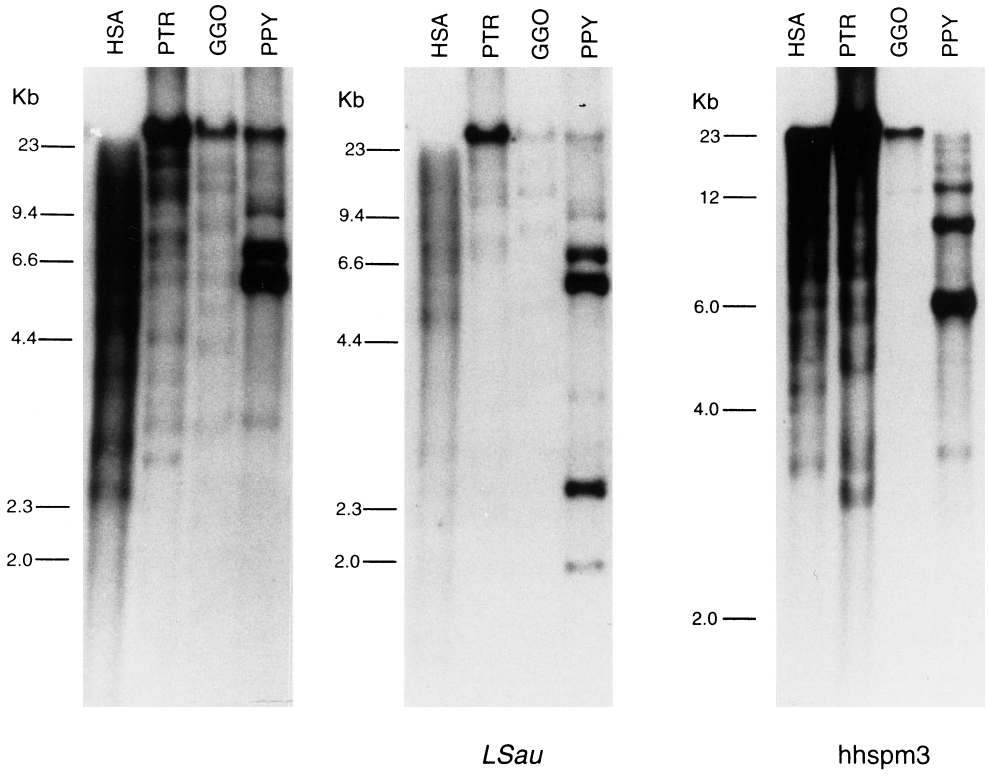

Fig. 3. Southern blots containing 8 µg *Eco*RI-digested genomic DNA from human (*HSA*), chimpanzee (*PTR*), gorilla (*GGO*) and orang-utan (*PPY*) were hybridised with probes to the D4Z4 repeat DNA motifs as indicated: homeobox, 9B6A; LSau, B6P1; hhspm3, BS1. The filters were washed in 2×SSC, 0.1% SDS at 65° C and exposed for 1–3 days

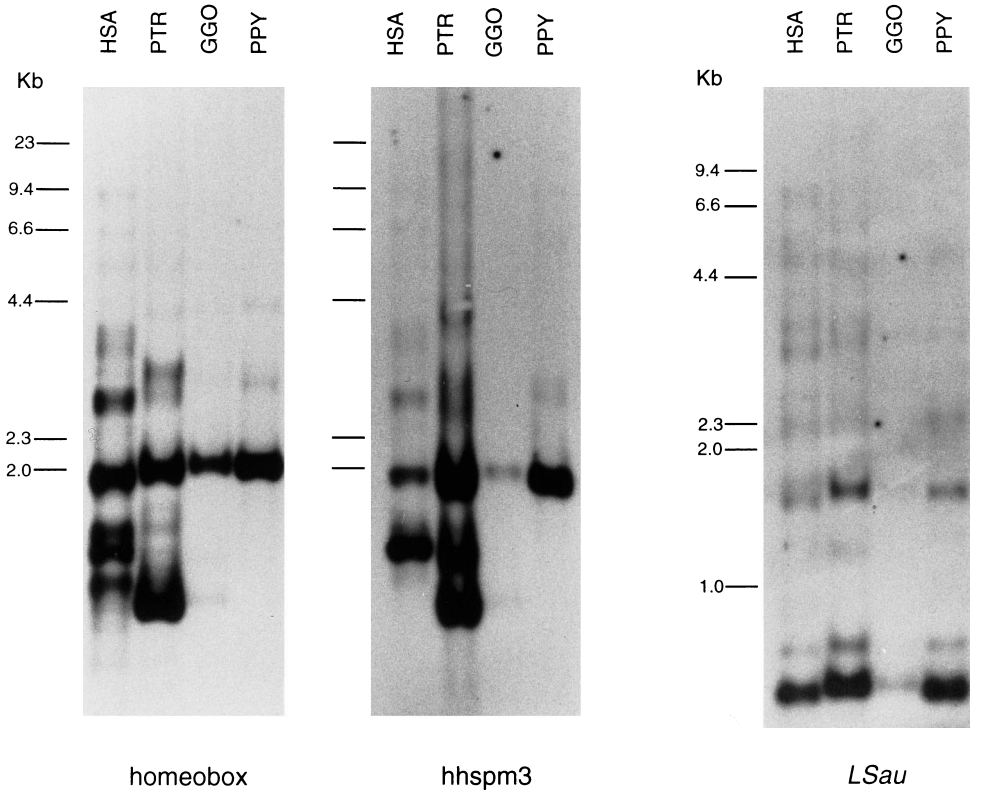

Fig. 4. Southern blots containing 8 µg *Pst*I-digested genomic DNA from human (*HSA*), chimpanzee (*PTR*), gorilla (*GGO*) and orang-utan (*PPY*) were hybridised with probes to the D4Z4 repeat DNA motifs as indicated: homeobox, 9B6A; hhspm3, BS1; LSau, B6P1. The filters were washed in 2×SSC, 0.1% SDS at 65° C and exposed for 3–4 days

electrophoresis (PFGE) is required to visualise all the alleles (Wijmenga et al. 1994). As the D4Z4 probes hybridised to *Eco*RI fragments larger than 23 kb in the great apes we used PFGE of *Eco*RI-digested high molecular weight genomic DNA to investigate these fragments in more detail. Probes to the homeobox, LSau and hhspm3 motifs all identified the same *Eco*RI fragments of 50–450 kb in human, chimpanzee and gorilla (Fig. 5). Taken together with the *Kpn*I data, this suggests the presence of similar tandem repeats in all these species.

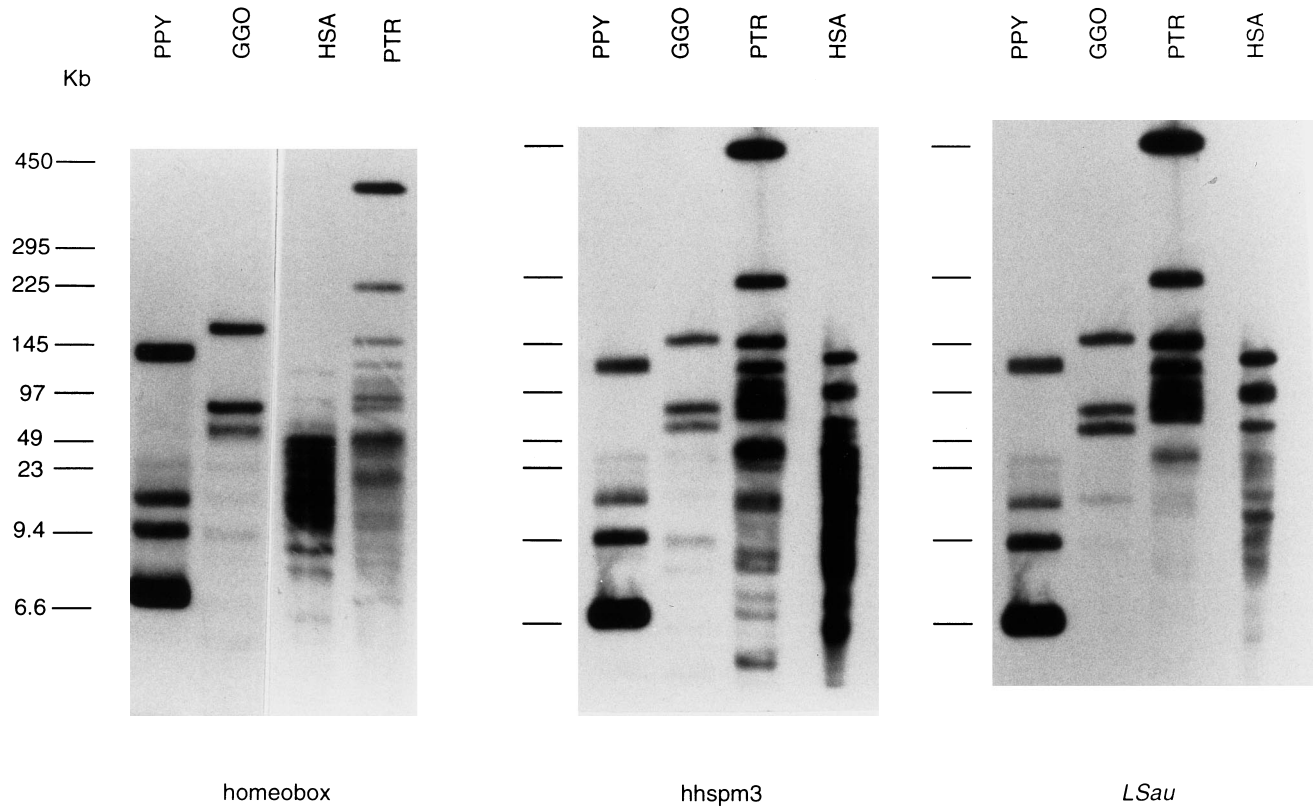

Fig. 5. Pulsed-field gel electrophoresis (PFGE) of genomic DNA from great apes. High molecular weight DNA embedded in agarose blocks was digested with *Eco*RI. The DNA fragments were separated by PFGE using a switch time of 3–30 s at 6 V/cm for 18 h. Following transfer of the DNA to nylon membranes, the filters were hybridised with probes to the D4Z4 repeat DNA motifs as indicated: homeobox, 9B6A; hhspm3, BS1; LSau, B6P1. The filters were washed in 1×SSC at 65° C and exposed for 2–3 days

In orang-utan, only one *Eco*RI fragment greater than 50 kb was present with many strongly hybridising bands in the 6–20 kb size range, consistent with the previous Southern blot data. Therefore, although orang-utan apparently has a 3.3-kb *Kpn*I repeat unit (Fig. 2), the long-range organisation may be different.

Evidence for a D4Z4 homologue in rhesus macaque

We have shown previously that the homeobox probe 9B6A produces a much simpler hybridisation pattern on Southern blots of DNA from Old and New World monkeys than in the great apes or man (Hewitt et al. 1994). Therefore, the conservation of the D4Z4 repeat in rhesus macaque, an Old World monkey, was investigated in more detail by Southern blotting and PFGE (Figs. 6, 7). The homeobox probe detected *Eco*RI fragments >23 kb, and *Pst*I, *Hind*III and *Bam*HI fragments of 6–6.5 kb. A similar hybridisation pattern was seen with this probe in another species of Old World monkey, crab-eating macaque (Fig. 6), with *Eco*RI and *Kpn*I fragments of >23 kb. In rhesus macaque, the LSau probe gave a similar hybridisation pattern to 9B6A, suggesting that these two sequences are also closely associated in this species. The hhspm3 probe also gave a similar hybridisation pattern, although several additional bands were seen. Thus, all the sequence elements within D4Z4 are conserved and map to fragments of similar sizes, suggesting the presence of D4Z4-like loci in Old World monkeys.

Again, we used PFGE to resolve accurately the sizes of the *Eco*RI fragments in two unrelated rhesus macaque individuals (Fig. 7). An identical hybridisation pattern was produced with each of the D4Z4 probes, again consistent with the presence of a similar repeat to D4Z4 in rhesus. Unlike the great apes, where *Kpn*I cuts within the repeat unit (Fig. 2), in rhesus *Eco*RI and *Kpn*I gave a identical hybridisation pattern with 9B6A (Fig. 7). In human *Hind*III cuts outside D4Z4, producing a similar hybridisation pattern to *Eco*RI (Wijmenga et al. 1994). We found that *Hind*III digestion produced a similar hybridisation pattern to *Eco*RI in gorilla and orang-utan (Fig. 7). However, in rhesus macaque no large fragments were seen, but a strongly hybridising band of 6.5 kb. This is consistent with a tandem repeat locus in rhesus with a repeat unit size of approximately 6–7 kb, rather than the 3.3 kb repeat seen in man and great apes.

Two unrelated macaque individuals were studied to examine whether these large *Eco*RI fragments were polymorphic. In both animals the D4Z4 probes identified up to three fragments of different sizes, indicating that rhesus macaque contains more than one polymorphic locus. It is noteworthy that it is often possible to resolve only three out of the possible four polymorphic

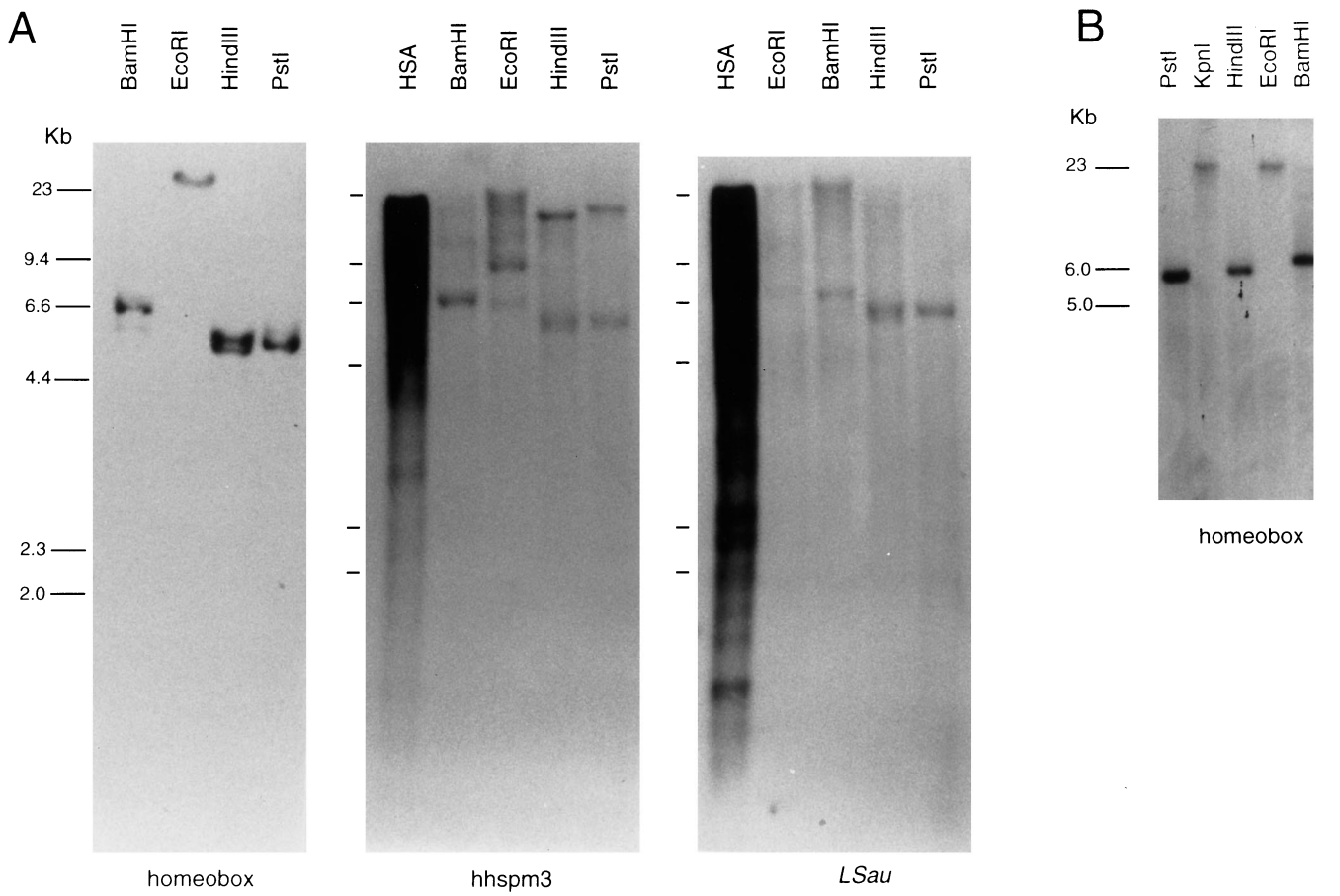

Fig. 6A, B. Southern blots containing 10 µg genomic DNA from rhesus macaque (**A**) or crab-eating macaque (**B**), digested with *Eco*RI, *Pst*I, *Bam*HI, *Kpn*I or *Hind*III, were hybridised with probes to the D4Z4 repeat DNA motifs as indicated: homeobox, 9B6A; hhspm3, BS1; LSau, B6P1. The lane labelled *HSA* contains human genomic DNA digested with *Eco*RI as a positive control. The filters were washed in 2×SSC, 0.1% SDS at 65° C and exposed for 3 days

fragments on human PFGE blots using probes to the D4Z4 locus (Wijmenga et al. 1994; Lyle et al. 1995). The *Hind*III data in Fig. 7 are from two unrelated gorilla individuals and again, the results show the D4Z4 homologue to be polymorphic.

Localisation of D4Z4-related loci in the great apes and Old World monkeys

Human cosmid and YAC clones containing the D4Z4 repeat were hybridised to metaphase spreads from human, chimpanzee, gorilla, orang-utan and an Old World monkey, the crab-eating macaque (Fig. 8). In all the great apes, D4Z4 probes hybridized to the telomeric region of the long arm of chromosome 3, the homologue of human chromosome 4, as identified either by chromosome painting or 4′,6-diamidino-2-phenylindole banding (Fig. 8A, C, D). These loci most likely represent the homologues of D4Z4.

In all these species additional signals were also detected, many of which localised to the short arms of the acrocentric chromosomes, a similar distribution pattern to that observed in human (cf. Fig. 8A). Co-localisation of D4Z4 probes with a probe to the rDNA genes was used to confirm the presence of signals on the acrocentric chromosomes (data not shown). The large block of heterochromatin present at human chromosome 1q is not conserved in the great apes, which probably accounts for the lack of FISH signals on chromosome 1 in these species. Together with the Southern blot and PFGE data, these results suggest that there are 3.3 kb repeats at the homologous region to human chromosome 4q35 and on at least some of the acrocentric chromosomes in each of the great apes studied. FISH signals were also occasionally seen at telomeric locations on a non-acrocentric chromosome, possibly corresponding to the homologue of the repeat locus of human chromosome 10.

In the Old World monkey, signals were seen at the telomere of the long arm of the homologue of human chromosome 4, and at the telomere of a second chromosomal pair (Fig. 8E). No other signals were detected, in contrast to the dispersed hybridisation pattern seen in the great apes. This data is consistent with the much simpler hybridisation pattern seen on Southern blots and with the PFGE data.

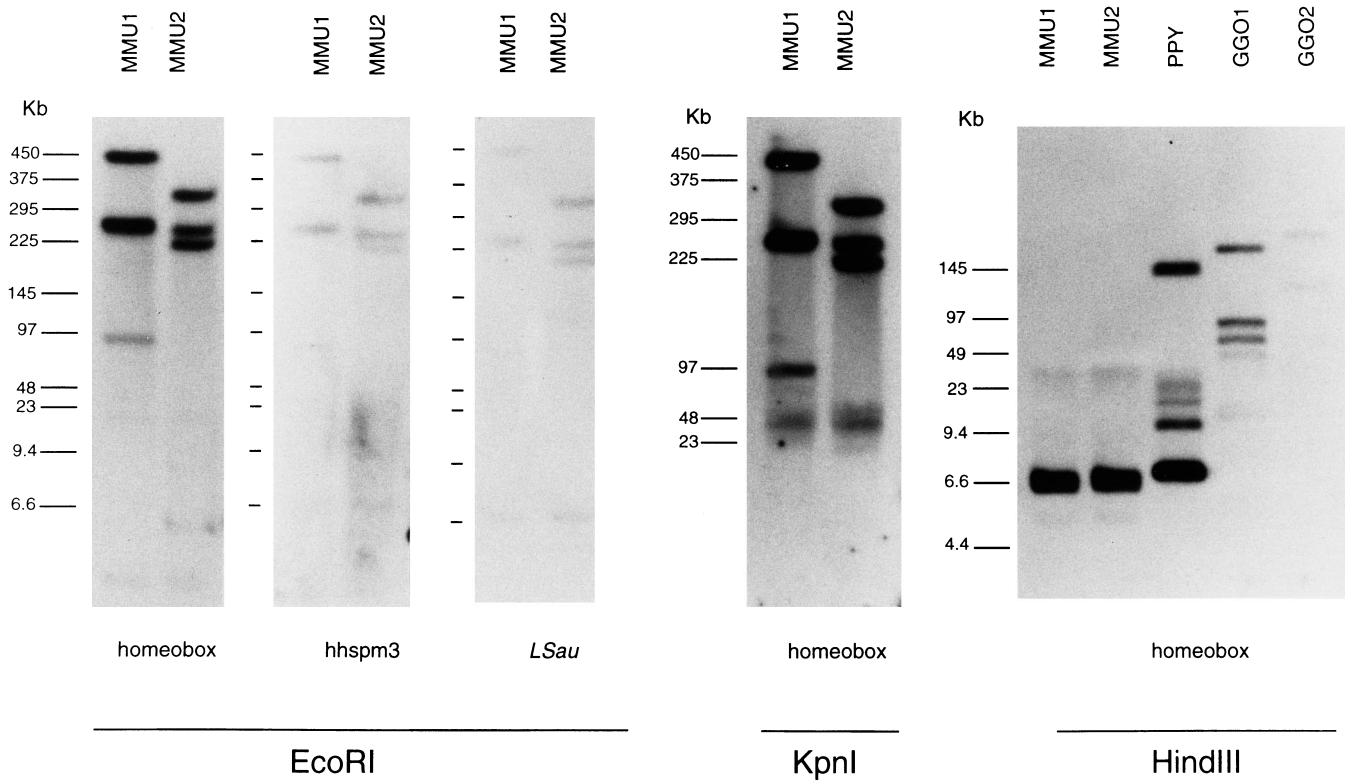

Fig. 7. PFGE of genomic DNA of rhesus macaque. High molecular weight DNA embedded in agarose blocks from two unrelated rhesus macaque individuals (*MMU1* and *MMU2*) was digested with *Eco*RI, *Kpn*I or *Hind*III. For the *Hind*III filters DNA from orang-utan (*PPY*) and two unrelated gorilla individuals (*GGO1* and *GGO2*) was also used. For *Eco*RI and *Kpn*I digests the DNA fragments were separated by PFGE using a switch time of 3–30 s at 6 V/cm for 18 h. For *Hind*III a switch time of 10–40 s at 5 V/cm for 24 h was used. Following transfer of the DNA to nylon membranes, the filters were hybridised with probes to the D4Z4 repeat DNA motifs as indicated: homeobox, 9B6A; hhspm3, BS1; LSau, B6P1. The filters were washed in 2×SSC at 65° C and exposed for 2–3 days

Discussion

We provide evidence that the D4Z4 repeat is conserved in great apes and Old World monkeys. Chimpanzee and man, in particular, gave an extremely similar pattern of hybridisation with all the probes, suggesting that the copy number and organisation of D4Z4 and other members of the 3.3 kb repeat family are similar in these species. This is consistent with data from many studies that support the chimpanzee-human clade (Caccone and Powell 1989; Sibley et al. 1990). It is also in agreement with Meneveri et al. (1995), who suggested that the LSau sequence organisation in similar in man and chimpanzee. However, some differences were seen with the other great apes. In orang-utan, two strongly hybridising *Eco*RI fragments of approximately 6.0 and 7.5 kb were consistently identified by D4Z4 probes, suggesting that this species contains a locus closely related to D4Z4, but with internal *Eco*RI sites. In gorilla fewer, fainter, hybridising bands were consistently seen using both standard Southern and PFGE analysis and using two sources of DNA. This may be due to the presence of fewer loci or to a greater level of sequence divergence. Again, our findings are consistent with those of Meneveri et al. (1995), who demonstrated differences in copy number and organisation of 68-bp satellite DNA and LSau sequences in gorilla and orang-utan using Southern blot analysis.

The data from Old World monkeys are consistent with the presence of only two loci related to D4Z4, in contrast to the large number in great apes. Using FISH, we showed one of these to map to a region homologous to human chromosome 4q35 in crab-eating macaque. We speculate that the other FISH signal could represent the homologue of the locus on human chromosome 10q26 (Bakker et al. 1995; Deidda et al. 1995). Again, there appears to be a tandem array, although the repeat unit size may be different.

Thus, great apes and Old World monkeys contain D4Z4-like loci at regions syntenic with human chromosome 4q35. In all great apes there are 3.3 kb repeat sequences on at least a subset of the acrocentric chromosomes, similar to the distribution of this family in the human genome (Hewitt et al. 1994; Winokur et al. 1994; Lyle et al. 1995). Therefore, we propose that there has been an increase in copy number of D4Z4-related sequences from a presumptive ancestral sequence through two pathways: (i) production of a tandem array and duplication of this locus onto another chromosome; (ii) dispersion of 3.3 kb repeat sequences into regions of heterochromatin, such as the short arms of the acrocentric chromosomes.

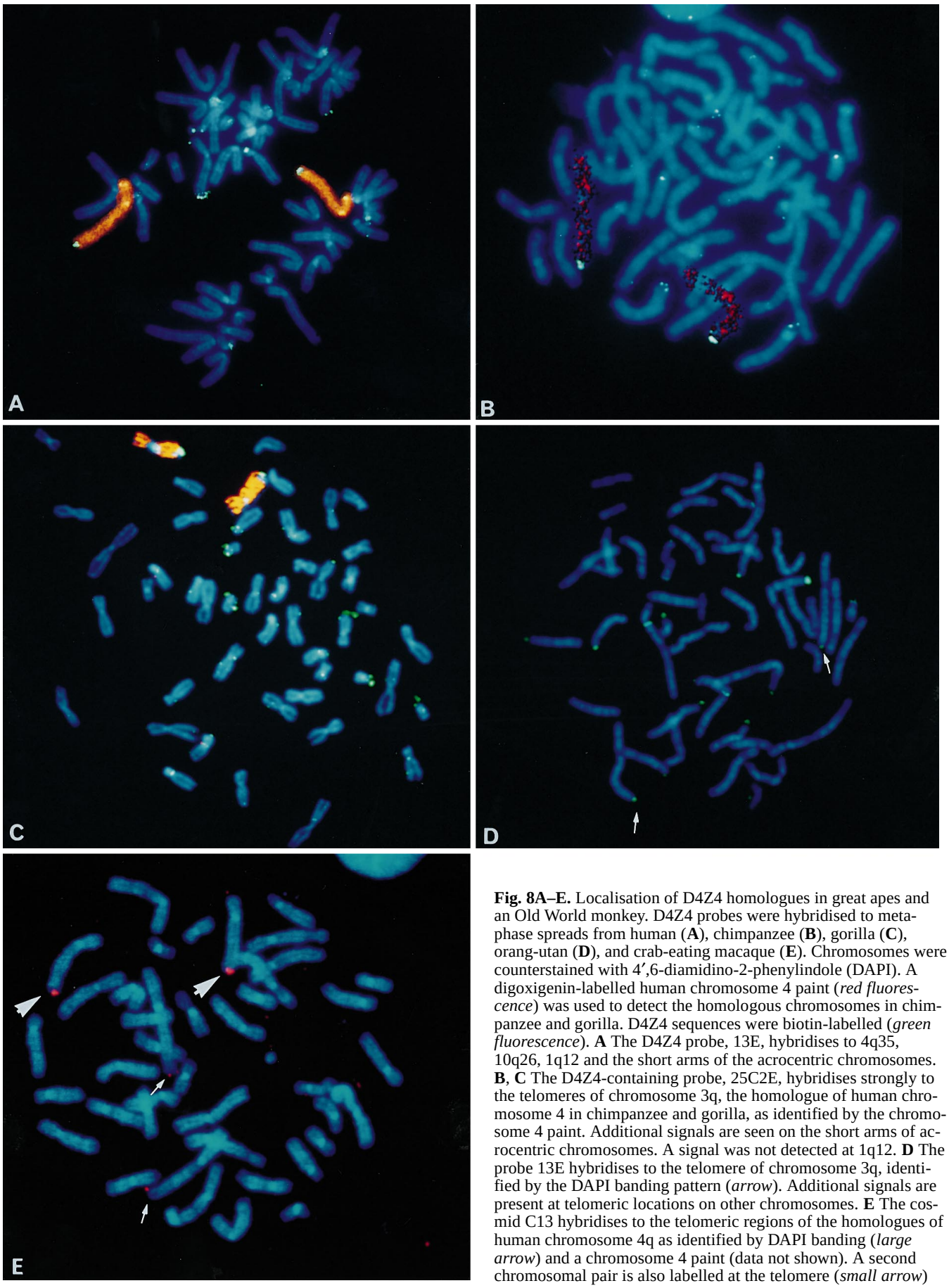

Fig. 8A–E. Localisation of D4Z4 homologues in great apes and an Old World monkey. D4Z4 probes were hybridised to metaphase spreads from human (**A**), chimpanzee (**B**), gorilla (**C**), orang-utan (**D**), and crab-eating macaque (**E**). Chromosomes were counterstained with 4′,6-diamidino-2-phenylindole (DAPI). A digoxigenin-labelled human chromosome 4 paint (*red fluorescence*) was used to detect the homologous chromosomes in chimpanzee and gorilla. D4Z4 sequences were biotin-labelled (*green fluorescence*). **A** The D4Z4 probe, 13E, hybridises to 4q35, 10q26, 1q12 and the short arms of the acrocentric chromosomes. **B**, **C** The D4Z4-containing probe, 25C2E, hybridises strongly to the telomeres of chromosome 3q, the homologue of human chromosome 4 in chimpanzee and gorilla, as identified by the chromosome 4 paint. Additional signals are seen on the short arms of acrocentric chromosomes. A signal was not detected at 1q12. **D** The probe 13E hybridises to the telomere of chromosome 3q, identified by the DAPI banding pattern (*arrow*). Additional signals are present at telomeric locations on other chromosomes. **E** The cosmid C13 hybridises to the telomeric regions of the homologues of human chromosome 4q as identified by DAPI banding (*large arrow*) and a chromosome 4 paint (data not shown). A second chromosomal pair is also labelled at the telomere (*small arrow*)

Generation of tandem arrays

A possible mechanism for the generation of a tandem array involves duplication from a progenitor sequence with subsequent recombination processes, such as unequal crossover and sister chromatid exchange, leading to the formation of tandem arrays of repeats. Similar mechanisms may be involved in producing the polymorphisms and the FSHD-associated deletions seen within D4Z4 (van Deutekom et al. 1993). This tandem array was then duplicated onto another chromosome, perhaps via a translocation event. The data presented here are consistent with either of the loci on chromosome 4 or chromosome 10 representing the progenitor locus. The homogeneity of the D4Z4 tandem array does not in itself argue against this repeat containing coding sequence; tandemly repeated gene families are characteristically homogeneous and intergenic regions are no less homogeneous than coding regions (Pavelitz et al. 1995).

Our data suggest that all the primate species studied here contain tandem arrays of D4Z4-related sequences. Rhesus macaque has an apparent repeat unit that is much larger than that found in the apes; cloning and sequence of these loci will be necessary to determine the molecular basis of this difference. The apparent homogeneity of the repeat units within a species presumably arose through concerted evolution, including such processes as gene conversion and unequal crossover.

Dispersion of the 3.3 kb repeat family

The results presented here suggest that the dispersion of 3.3 kb repeat loci onto the acrocentric chromosomes occurred after the split of the great ape and Old World monkey lineages. Although this is based on data obtained mainly from two species of Old World monkey (rhesus macaque and crab-eating macaque), Southern blot analysis in baboon and a New World monkey (marmoset) is consistent with dispersion occurring after the divergence of the great apes and Old World monkeys (this paper and Hewitt et al. 1994).

Given the distribution of the 3.3 kb repeat family in human and great ape genomes and its association with heterochromatic regions, it seems likely that sequence exchanges between heterochromatic regions of the acrocentric chromosomes may have facilitated the evolution of this repeat family. Interestingly, we have found that the heterochromatin-associated (68 bp satellite, which is associated with the D4Z4 repeat (van Deutekom et al. 1993; Hewitt et al. 1994) and other members of the 3.3 kb repeat family (Lyle et al. 1995) hybridised only to DNA from man or the great apes and not to rhesus macaque (data not shown). An expansion of this satellite DNA has occurred in man; this is thought to be related to the ability of this repeat to become excised and circularised and/or to undergo frequent unequal crossovers (Assum et al. 1993). Therefore, this association of the D4Z4 repeat with β satellite DNA may have contributed to its dispersion.

Other primate-specific repetitive sequence families that map to heterochromatic regions include the 724 family (Kurnit et al. 1986) and the multisequence family chAB4 (Assum et al. 1994). Both these families have also undergone rapid expansion recently during primate evolution and it has been suggested that sequence exchanges between heterochromatic regions are relatively frequent and play an important role in the dispersion of repetitive DNA (Assum et al. 1994). There is good evidence that sequence exchanges between heterochromatic regions of the genome occur. Several repetitive DNA families are shared by the acrocentric chromosomes, such as the subfamilies of alphoid DNA (Choo et al. 1988). Similar mechanisms have probably resulted in distribution of the 3.3 kb repeat family as we have shown previously that subfamilies of the 3.3 kb repeat are shared by more than one chromosome (Lyle et al. 1995). A more detailed molecular analysis of D4Z4 and its homologues in other primate species, including cloning of DNA sequences flanking the tandem array, should give insights into the molecular evolution of this locus. In addition analysis of more divergent primate species will be important. Given the large sizes of these repeat regions, the availability of resources (such as YACs and bacterial artificial chromosomes) for cloning of large pieces of DNA from primate species would be extremely advantageous.

Implications for identification of the FSHD gene

The finding that tandem arrays closely related to D4Z4 have been maintained at loci homologous to human chromosome 4q35–qter in apes and Old World monkeys suggests a functionally important role for these sequences. The D4Z4 locus maps within 50 kb of the telomere of human chromosome 4q (Bengtsson et al. 1994; Hewitt et al. unpublished data). Whether the telomeric localisation of the repeat has played a significant role in its evolution is unknown. To date, the mechanism whereby rearrangements within the D4Z4 repeat sequence cause FSHD is still unresolved. The presence of many copes of the 3.3 kb repeat in man has complicated the study of the function of this locus (Lyle et al. 1995). The data presented here suggest that Old World monkeys, in which the tandem array has been maintained but that lack the many copies of similar sequences, may provide a good model for functional analysis of the repeat. In addition, human chromosome 4q35–qter contains many mildly repetitive sequences and pseudogenes that have proved problematic in gene-searching strategies (Altherr et al. 1995; van Deutekom et al. 1995; Hewitt et al. manuscript in preparation). Cloning of the homologous region in rhesus macaque may provide a less complex resource in which to identify bona fide genes that can then be tested as candidate genes for FSHD.

Acknowledgements. The authors wish to thank Robert Lyle for helpful and entertaining discussions, Tracy Wright for cosmid C13, Dr. D. Hunt for providing primate tissue samples and Dr. Charles Streuli for generously allowing the use of his tissue culture facilities. This work was funded by The Wellcome Trust, the MDA (USA) and the Royal Society. L. N. C. was a recipient of a HUGO Travel Award.

References

Agresti A, Meneveri R, Siccardi A, Marozzi A, Corneo G, Gaudi S, Ginelli E (1989) Linkage in human heterochromatin between highly divergent *Sau*3A repeats and a new family of repeated DNA sequences (*Hae*III family). J Mol Biol 205: 625–631

Altherr MR, Bengtsson U, Markovich RP, Winokur ST (1995) Efforts toward understanding the molecular basis of facioscapulohumeral dystrophy. Muscle Nerve Suppl 2:S32–S38

Assum G, Fink T, Steinbeißer T, Fisel K (1993) Analysis of human extrachromosomal DNA elements originating from different β-satellite subfamilies. Hum Genet 91:489–495

Assum G, Gartmann C, Schempp W, Wohr G (1994) Evolution of the chAB4 multisequence family in primates. Genomics 21:34–41

Bakker E, Wijmenga C, Vossen R, Padberg GW, Hewitt JE, Vanderwielen M, Rasmussen K, Frants RR (1995) The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a homologue on 10qter. Muscle Nerve Suppl 2:S39–S44

Bengtsson U, Altherr MR, Wasmuth JJ, Winokur ST (1994) High resolution fluorescence *in situ* hybridisation to linearly extended DNA visually maps a tandem repeat associated with facioscapulohumeral muscular dystrophy immediately adjacent to the telomere of 4q. Hum Mol Genet 3:1801–1805

Caccone A, Powell JR (1989) DNA divergence among hominoids. Evolution 43:925–942

Choo KH, Vissel B, Brown R, Filby RG, Earle E (1988) Homologous alpha satellite sequences on human acrocentric chromosomes with selectivity for chromosomes 13, 14 and 21: implications for recombination between nonhomologous and Robertsonian translocations. Nucleic Acids Res 16:1273–1284

Church GM, Gilbert W (1984) Genomic sequencing. Proc Natl Acad Sci USA 81:1991–1995

Deidda G, Cacurri S, Grisanti P, Vigneti E, Piazzo N, Felicetti L (1995) Physical mapping evidence for a duplicated region on chromosome 10qter showing high homology with the FSHD locus on chromosome 4qter. Eur J Hum Genet 3:155–167

Hewitt JE, Lyle R, Clark LN, Valleley EM, Wright TJ, Wijmenga C, van Deutekom JCT, Francis F, Sharpe PT, Hofker M, Frants RR, Williamson R (1994) Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. Hum Mol Genet 3:1287–1295

Jauch A, Wienberg J, Stanyon R, Arnold N, Tofanelli S, Ishida T, Cremer T (1992) Reconstruction of genomic rearrangements in great apes and gibbons by chromosome painting. Proc Natl Acad Sci USA 89:8611–8615

Kurnit DM, Roy S, Stewart GD, Schwedock J, Neve RL, Bruns GAP, Van Keuren MI, Patterson D (1986) The 724 family of DNA sequences is interspersed about the pericentromeric regions of human acrocentric chromosomes. Cytogenet Cell Genet 43:109–116

Lunt PW, Harper PS (1991) Genetic counselling in facioscapulohumeral muscular dystrophy. J Med Genet 28:665–664

Lyle R, Wright TJ, Clark LN, Hewitt JE (1995) The FSHD-associated repeat, D4Z4, is a member of a dispersed family of homeobox-containing repeats, subsets of which are clustered on the short arms of the acrocentric chromosomes. Genomics 28:389–397

Meneveri R, Agresti A, Marozzi A, Saccone S, Rocchi M, Archidiacono N, Corneo G, Valle GD, Ginelli E (1993) Molecular organisation and chromosomal location of human GC-rich heterochromatic blocks. Gene 123:227–234

Meneveri R, Agresti A, Marozzi A, Ginelli E (1995) Analysis of GC-rich repetitive nucleotide sequences in great apes. J Mol Evol 40:405–412

Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16:1215

Passos-Bueno MR, Wijmenga C, Takata RE, Marie SKN, Vainzof M, Pavanello RC, Hewitt JE, Bakker E, Carvalho A, Akiyama J, Frants RR, Zatz M (1993) No evidence of genetic heterogeneity in Brazilian facioscapulohumeral muscular dystrophy families (FSHD) with 4q markers. Hum Mol Genet 2:557–562

Pavelitz T, Rusché L, Matera AG, Scharf JM, Weiner AM (1995) Concerted evolution of the tandem array encoding primate U2 snRNA occurs *in situ*, without changing the cytological context of the *RNU2* locus. EMBO J 14:169–177

Ried T, Arnold N, Ward DC, Wienberg J (1993) Comparative high-resolution mapping of human and primate chromosomes by fluorescence *in situ* hybridization. Genomics 18:381–386

Sibley C, Comstock J, Ahlquist J (1990) DNA hybridisation evidence of hominoid phylogeny: a reanalysis of the data. J Mol Evol 30:202–236

Tyler-Smith C, Willard HF (1993) Mammalian chromosome structure. Curr Opin Genet Dev 3:390–397

Upadhyaya M, Jardine P, Maynard J, Farnham J, Sarfarazi M, Wijmenga C, Hewitt JE, Frants R, Harper PS, Lunt PW (1993) Molecular analysis of British facioscapulohumeral dystrophy families of 4q DNA rearrangements. Hum Mol Genet 2:981–987

van Deutekom JCT, Wijmenga C, van Tienhoven EAE, Gruter A-M, Hewitt JE, Padberg GW, van Ommen G-JB, Hofker MH, Frants RR (1993) FSHD associated rearrangements are due to deletion of integral copies of a 3.2 kb tandemly repeated unit. Hum Mol Genet 2:2037–2042

van Deutekom JCY, Hofker MH, Romberg S, van Geel M, Rommens J, Wright TJ, Hewitt JE, Padberg GW, Wijmenga C, Frants RR (1995) Search for the FSHD gene using cDNA selection in a region spanning 100 kb on chromosome 4q35. Muscle Nerve Suppl 2:S19–S26

van Deutekom J, Lemmers R, Grewal P, van Geel M, Romberg S, Dauwerse H, Wright T, Padberg G, Hofker M, Hewitt J, Frants R (1996) Identification of the first gene (FRG1) from the FSHD region on human chromosome 4q35. Hum Mol Genet 5:581–589

Weiffenbach B, Dubois J, Storvick D, Tawil R, Jacobsen SJ, Gilbert J, Wijmenga C, Mendell JR, Winokur S, Altherr MR, Schultz P, Olandt S, Frants RR, Pericak-Vance M, Griggs RC (1993) Mapping the facioscapulohumeral muscular dystrophy gene is complicated by chromosome 4q35 recombination events. Nat Genet 4:165–169

Wijmenga C, Hewitt JE, Sandkuijl LA, Clark LN, Wright TJ, Dauwerse HG, Gruter A-M, Hofker MH, Moerer P, Williamson R, van Ommen G-JB, Padberg GW, Frants RR (1992) Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. Nat Genet 2:26–30

Wijmenga C, van Deutekom JCT, Hewitt JE, Padberg GW, van Ommen G-JB, Hofker MH, Frants RR (1994) Pulsed-field gel electrophoresis of the D4F104S1 locus reveals the size and parental origin of the facioscapulohumeral muscular dystrophy (FSHD)-associated deletions. Genomics 19:21–26

Winokur ST, Bengtsson U, Feddersen J, Mathews KD, Weiffenbach B, Bailey H, Markovich RP, Murray JC, Wasmuth JJ, Altherr MR, Schutte BC (1994) The DNA rearrangement associated with facioscapulohumeral muscular dystrophy involves a heterochromatin-associated repetitive element: implications for a role of chromatin structure in the pathogenesis of the disease. Chromosome Res 2:225–234

Wright TJ, Wijmenga C, Clark LC, Frants RR, Williamson R, Hewitt JE (1993) Fine mapping of the FSHD gene region orientates the rearranged fragment detected by the probe p13E-11. Hum Mol Genet 2:1673–1678

Zhang X-Y, Loflin PT, Gehrke CW, Andrews PA, Ehrlich M (1987) Hypermethylation of human DNA sequences in embryonic carcinoma cells and somatic tissues but not in sperm. Nucleic Acids Res 15:9429–9449