Evolutionary history of linked D4Z4 and Beta satellite clusters at the FSHD locus (4q35)☆

Marta Giussani [a,1], Maria Francesca Cardone [b,2], Beatrice Bodega [a,3],
Enrico Ginelli [a], Raffaella Meneveri [c,*]

[a] Department of Biology and Genetics for Medical Sciences, University of Milan, Milan, Italy
[b] Department of Genetics and Microbiology, University of Bari, Bari, Italy
[c] Department of Experimental Medicine, University of Milano-Bicocca, Monza, Italy

A B S T R A C T

We performed a detailed genomic investigation of the chimpanzee locus syntenic to human chromosome 4q35.2, associated to the facioscapulohumeral dystrophy. Two contigs of approximately 150 kb and 200 kb were derived from PTR chromosomes 4q35 and 3p12, respectively: both regions showed a very similar sequence organization, including D4Z4 and Beta satellite linked clusters. Starting from these findings, we derived a hypothetical evolutionary history of human 4q35, 10q26 and 3p12 chromosome regions focusing on the D4Z4–Beta satellite linked organization. The D4Z4 unit showed an open reading frame (DUX4) at both PTR 4q35 and 3p12 regions; furthermore some subregions of the Beta satellite unit showed a high degree of conservation between chimpanzee and humans. In conclusion, this paper provides evidence that at the 4q subtelomere the linkage between D4Z4 and Beta satellite arrays is a feature that appeared late during evolution and is conserved between chimpanzee and humans.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

The human genome contains a large amount (>50%) of repetitive sequences, which were, for a long time, regarded as parasitic or "junk" DNA with no phenotypic impact [1,2]. However, recent data suggest that retroelements as well as tandem repeats (i.e. satellite DNA) are involved in mammalian gene regulation [3,4]. In fact, transcripts derived from satellite DNAs seem to participate in the epigenetic process(es) of chromatin remodeling and heterochromatin formation, which are crucial for genome stability [5,6]. Furthermore, some satellite DNAs show the conservation of particular structural motifs; these selective constraints are probably related to their interaction with proteins involved in heterochromatin formation [7]. An example of the importance

of clustered repeats for human cell physiology can be derived from the contraction of the 3.3 kb (D4Z4) megasatellite located at chromosome 4q35.2 which is associated to FSHD manifestation [8–10]. FSHD is an autosomal dominant disease showing at the subtelomere of chromosome 4q 1–10 copies of D4Z4, whereas in the normal population the number of repeats is highly polymorphic ranging from 11 to 150. The 4q subtelomere comes in two genomic variants [11], alleles 4qA and 4qB, which are equally common in the general population and show an equal variation in size and somatic instability [12]. Nonetheless, FSHD alleles are always of the 4qA type [13], which differs from 4qB due to the presence, downstream to the D4Z4 array, of the pLAM sequence (260 bp) followed by the Beta satellite cluster of 6.2 kb in length. In normal individuals the FSHD region is methylated and displays the typical feature of unexpressed euchromatin [14,15], whereas in FSHD patients hypomethylation of the contracted D4Z4 alleles has been reported [16–18]. Furthermore, DNA hypermethylation is associated with the recruitment of various methylated DNA binding proteins as well as histone-modifying deacetylases and methyltransferases, all involved in chromatin condensation [19]. Thus, in FSHD hypomethylation of the contracted 4q D4Z4 alleles supports the relevance of these epigenetic changes able to induce a more euchromatic-like D4Z4 array, a fact that seems directly associated with FSHD expression [20]. Both histone modifications and DNA hypomethylation of D4Z4 are likely to be disease-related, affecting in some way the functionality of the 4qter region, most probably through the up-regulation of D4Z4-associated genes. Accordingly, it has been observed that the number of residual

D4Z4 copies is inversely correlated to the age of onset and severity of the disease [21]. Nevertheless, patients showing the same extent of the deletion may differ in disease progression [22], and monozygotic twins show discordant or extremely variable phenotypes. Furthermore, a small percentage of FSHD patients (phenotypic FSHD or FSHD-2) do not carry the D4Z4 contraction at 4q35.2, whereas some individuals are asymptomatic carriers of the D4Z4 contraction [23,24]. These observations strongly suggest that FSHD manifestation might rely on epigenetic modifications other than those attributable to D4Z4.

To address this question, after having defined the molecular features of the 4q region in the gorilla [25], we decided to investigate the sequence composition of the orthologous region in chimpanzee (PTR). Since most biological phenomena can be better understood if studied in an evolutionary frame, the identification of conserved and/or lost structural differences between human and primate 4q subtelomeres might be relevant to understand the etiology of the disease. This paper reports that at the 4q subtelomere the linkage between D4Z4 and Beta satellite arrays is a feature that seems to be appeared late during evolution and is conserved between chimpanzee and humans, and that Beta satellite sequences may play a role in the epigenetic regulation of the 4q35.2 region.

2. Materials and methods

2.1. Genomic library screening

The genomic library was obtained from BAC/PAC Resources, Children's Hospital, Oakland Research Institute (CHORI-251 segments 3 and 4: Chimpanzee BAC Library). High-density arrayed BAC filters were hybridized according to the instructions provided by BAC/PAC Resources [BACPAC Resources Center [http://bacpac.chori.org/]] with a ^{32}P-labeled D4Z4 probe (77 M12-KpnI D4Z4 probe) obtained by subcloning a KpnI fragment of 3.3 kb from a previously isolated human genomic BAC (77 M12) from the human shared BAC library CHORI-16 (BACPAC Resources Center).

2.2. DNA isolation and Southern blot hybridization

BAC clones were grown in LB medium supplemented with chloramphenicol. BAC DNA was purified using a Sigma PhasePrep BAC kit (Sigma Aldrich, USA), digested with restriction enzymes (Biolabs, USA), fractionated by means of agarose gel electrophoresis, blotted onto nylon filters (Hybond N+, Amersham, UK), and hybridized with a ^{32}P-labeled probe representative of D4Z4 (77 M12-KpnI D4Z4) and Beta satellite sequences (A17a plasmid) [26].

Molecular hybridization was performed in 2× SSC at 60 °C overnight, and the filters were washed at 60 °C in 1× SSC twice for 20 min. The hybridization signals were quantified by means of phosphoimaging (Typhoon 9200, Amersham).

2.3. DNA sequencing and in silico analysis

T7/Sp6 BAC ends were sequenced using the Big Dye terminator 3.1 system as described by BAC/PAC Resources [BACPAC Resources Center [http://bacpac.chori.org/]]. All fluorescent traces were analyzed using the Applied Biosystem Model 3100 DNA Sequencing System (Applied Biosystem). The newly derived nucleotide sequences are available at NCBI [http://www.ncbi.nlm.nih.gov/] with accession numbers: JM170483 (584M21-Sp6), JM170484 (584M21-T7), JM170481 (578I19-Sp6), JM170482 (578I19-T7), JM170493 (709I17-Sp6), JM170494 (709I17-T7), JM170491 (705I9-Sp6), JM170492 (705I9-T7), JM170497 (784B11-Sp6), JM170498 (784B11-T7), JM170501 (802O1-Sp6), JM170502 (802O1-T7), JM170499 (790M3-Sp6), JM170500 (790M3-T7), JM170495 (773B17-Sp6), JM170496 (773B17-T7), JM170485 (590D11-Sp6), JM170486 (590D11-T7), JM170487 (618G2-Sp6), JM170488 (618G2-T7), JS886734 (678H18-Sp6), JS886735 (678H18-T7), JM170489 (700D23-Sp6), JM170490 (700D23-T7), JM170503 (809J1-Sp6), and JM170504 (809J1-T7). Internal BAC sequencing was carried out at BMR Genomics (Padua, Italy) by the ROCHE-GS-FLX 454 platform. A total of four BACs were sequenced and the newly derived nucleotide sequences are available at NCBI with accession numbers: JF900598 (BAC 584M21), JF900596 (BAC 709I17), JF900597 (BAC 705I9), and JF900595 (BAC 809J1). DNA sequence analysis was performed using DNASTAR software and NCBI and UCSC Genome Bioinformatics Sites facilities [http://www.ncbi.nlm.nih.gov/; http://genome.cse.ucsc.edu/].

2.4. Fluorescent in situ hybridization (FISH) on metaphase chromosome spreads

The metaphase chromosome spreads were obtained by standard methods from *Pan troglodytes* (PTR) lymphoblastoid or fibroblast cell lines. The primate cell lines were provided by M. Rocchi [27]. Probes were labeled with digoxigenin-dUTP or biotin-dUTP (Roche Diagnostic) using a nick translation kit (Roche Diagnostic), and detected either using fluorescein-streptavidin (SIGMA) for the biotin-labeled probes, or CY3-anti-digoxigenin conjugated antibodies (Jackson ImmunoResearch Laboratories) for the digoxigenin-labeled probes. The chromosomes were counterstained with propidium iodide and DAPI in antifade (Vectashild), and then visualized using a Leitz DM-RB microscope equipped with DAPI and FITC/TRITC epifluorescence optics. Hybridizations were performed in 50% formamide (v/v), 10% dextran sulfate, 2× SSC at 37 °C, in the presence of human Cot1 DNA (Gibco-BRL). Post-hybridization washing included 50% formamide, 2× SSC at 42 °C, followed by three washes in 1× SSC at 60 °C (HSA), or 50% formamide, 1× SSC at 37 °C, followed by three washes in 1× SSC at 42 °C (GGO and PTR). The chromosomes were counterstained with DAPI (4′,6-diamidino-2-phenylindole), and digital images were captured using a Leica DMRXA epifluorescence microscope equipped with a cooled CCD camera (Princeton Instruments). The fluorescence signals were recorded separately as gray-scale images. The images were pseudocolored and merged using Adobe Photoshop software.

In all the paper great apes' chromosomes are identified as the homologous human chromosomes.

3. Results

3.1. Isolation and characterization by Southern blot hybridization of D4Z4-positive PTR genomic clones

The PTR genomic library CHORI-251 (four genome equivalents) was screened with a human D4Z4 probe allowing the isolation of a total of thirteen BACs. The PTR clones were analyzed for the presence of D4Z4 and Beta satellite clusters by digestion with different restriction enzymes and hybridization with the corresponding human probes (Fig. S1). Four clones (584M21, 578I19, 709I17 and 705I9) were D4Z4+/Beta satellite+ (Fig. S1), and showed D4Z4 and Beta satellite restriction enzyme variants (Fig. S1 and Table S1). In particular, the clones 584M21/578I19 and 709I17/705I9 contained Kpn+/BlnI− and Kpn−/BlnI+ D4Z4 units, respectively. Furthermore, the clones 584M21/578I19 and 709I17/705I9 carried a Beta satellite cluster originating a comparable ladder of 68 bp mers when digested by Sau3A, but a distinct PstI restriction pattern (Fig. S1 and Supplementary Table 1). These results indicate that the two pairs of BACs 584M21/578I19 and 709I17/705I9 carry distinct D4Z4 and Beta satellite clusters, suggesting their derivation from different chimpanzee chromosome regions. Among the remaining nine clones, six (773B17, 790M3, 802O1, 809J1, 590D11, and 618G2) showed Kpn+/BlnI− D4Z4 units and three (784B11, 700D3 and 678H18) Kpn−/BlnI+ D4Z4 units (Fig. S1, Table S1 and not shown).

3.2. Terminal sequencing and databank analysis of D4Z4-positive PTR genomic clones

The isolated PTR clones were then characterized by end sequencing, and the derived sequences compared with the human genome databank (hg19). Considering D4Z4 +/Beta satellite + positive BACs, 584M21-Sp6 and 578I19-T7 showed a high percentage of identity with 4q35.2 (98.6% and 96.8%, respectively), whereas 584M21-T7 and 578I19-Sp6 with 3q29 (97.2% and 98.6%, respectively) (Table S2 and Fig. S2). Human genome databank analysis also evidenced that chromosome 3q29 contains duplicated sequences present in several subtelomeres, such as 1p36.33, 5q35.3 and 19q13.3 (not shown). Conversely, 709I17-T7 and 705I9-T7 showed identity with the 3p12.3 chromosome region, whereas the opposite ends aligned with 4q35.2 (709I17-Sp6) and with D4Z4 (705I9-Sp6) (Table S2 and Fig. S2). From these data it can be derived that BAC 578I19 is comprised within BAC 584 M21 (single asterisk in Fig. S2), and that 709I17 spans from the FRG2C gene on 3p12.3 to subtelomeric sequences located downstream to D4Z4/Beta satellite clusters on both 4q35 and 10q26 chromosomes (double asterisk in Fig. S2 and Table S2). Furthermore BAC 705I9 is delimited by D4Z4 and 3p12.3 sequences (triple asterisk in Fig. S2 and not shown). Since analyses by Southern blot hybridization showed the occurrence within these BACs of both D4Z4 and Beta satellite clusters, the results strongly indicate that in chimpanzee a very similar repetitive sequence block is located on both 4q35 and 3p12. Of the remaining nine clones, five showed at both ends D4Z4 sequences (809J1, 590D11, 618G2, 700D3 and 678H18), three D4Z4 and 4q35 sequences (773B17, 790M3 and 802O1), and one D4Z4 and 3p12 (784B11) sequences. Chromosome coordinates and identity, and GenBank accession numbers of the BAC ends are reported in Table S2 and in Fig. S2.

3.3. Chromosome location by FISH of D4Z4/Beta satellite positive PTR genomic clones

Total DNA or inter-Alu sequences from BACs 584M21, 578I19, 709I17 and 705I9 were used as probes on PTR metaphase chromosome spreads. As reported in Table S3 and in Fig. 1, the BACs can be subdivided into two classes, those essentially mapping to 4q35 (584M21 and 578I19) and those also highlighting the 3p12 chromosome region (709I17 and 705I9). These results provide additional evidence that, in chimpanzee, the 3p12 and 4q35 chromosome regions share a large duplication including sequences other than D4Z4 and Beta satellite.

3.4. Sequencing of D4Z4/Beta satellite positive PTR genomic clones

To define the sequence content and organization of 4q35 and 3p12 chromosomes in chimpanzee, BACs 584M21, 709I17 and 705I9 were sequenced. The sequencing project also included BAC 809J1 as representative of DNA inserts flanked at both ends by D4Z4 repeats. BAC inserts were not completely sequenced, and thus they contain some gaps of unknown length. The alignment with the human genome databank (hg19) evidenced for BAC 584M21 a high percentage of identity (>98%) with both 4q35.2 (subsequences 12, 17 and 18; total length of 52.888 kb) and 3q29 (subsequences 14 and 13; total length of 33.898 kb) (Table S4). Furthermore, in BAC 584M21, a fragment of D4Z4 linked to a cluster of Beta satellite of about 6 kb (subsequence 15 asterisked in Table S4; 11.488 kb in length) showed an identity of 96.2% with 10q26.3. Interestingly, this DNA region also showed a similar percentage of identity (>97%) with 18p11.32. Also the BAC 809J1 belonged to the 4q35 chromosome; in fact it aligned (>98% of identity) with the block of sequences that in human chromosome 4q35.2 are comprised
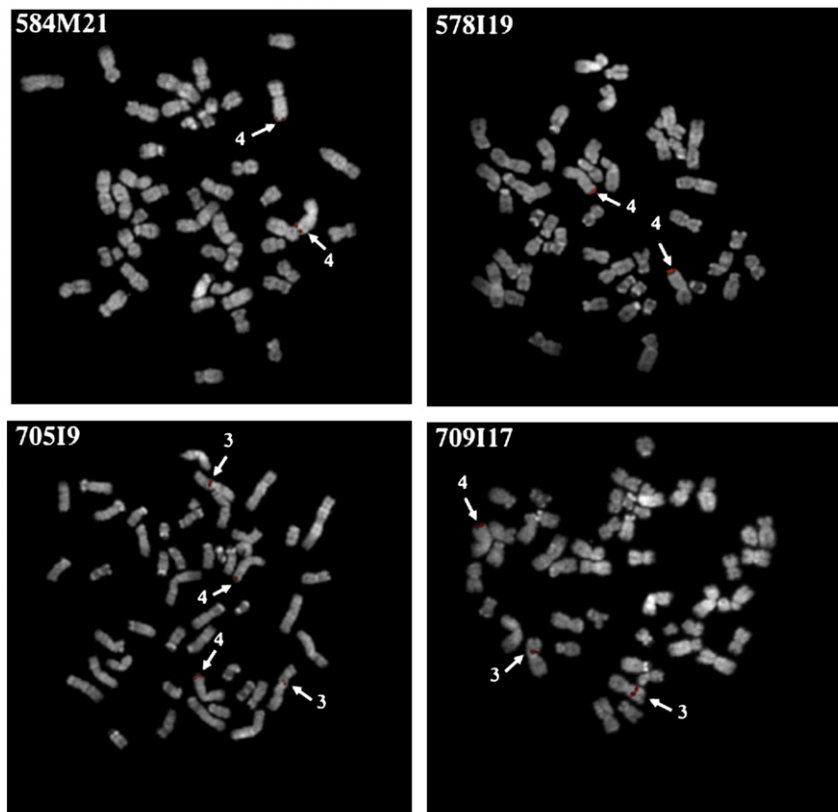

Fig. 1. Chromosomal location by FISH on chimpanzee metaphase chromosome spreads of unique and low repetitive sequences contained in PTR BACs 584M21, 578I19, 705I9 and 709I17. In all FISH experiments inter-Alu from the corresponding BAC was used as a probe. The roman numbers refer to the homologous human chromosome.

between DUX4c and the D4Z4 cluster (Table S4). Regarding the sequence organization of the 3p12 chromosome, it is noteworthy that BACs 709I17 (subsequence 35; 11.360 in length) and 705I9 (subsequence 69; 8.122 kb in length) (double asterisks in Table S4) overlapped with 99.8% of identity (not shown) and hence they were considered as a contig. The overlapping region showed the presence of a D4Z4–Beta satellite sequence organization very similar to that found in 4q35. The sequence upstream to the D4Z4/Beta satellite region (subsequences 36 and 37 of BAC 709I17; total length of 44.098 kb) displayed identity with both 3p12.3 and 4q35.2. In particular, 4q35.2-like sequences included the FRG2 gene and the polymorphic marker SSLP, whereas 3p12.3-like sequences a short cluster of divergent Beta satellite (approximately 2 kb in length) showing a high percentage of identity (98.4%) with the human orthologous region (not shown). Conversely, the sequence downstream to the D4Z4/Beta satellite region, (subsequences 67, 61 and 62 of BAC 705I9: total length of 46.066 kb) aligned only with 3p12.3 (identity index of approximately 97%). Coordinates on different chromosomes, identities with human genome databank (hg19) and accession numbers of the derived BAC sequences are listed in Table S4. A graphical representation of the alignment of subsequences from BACs 584M21, 709I17 and 705I9 on human chromosomes obtained by BLAT is reported in Figs. S3 and S4.

In summary, BAC 584M21 spans from upstream the FRG2 gene to subtelomeric sequences located downstream to the linked clusters of D4Z4 and Beta satellite on chromosomes 4q35.2/10q26.3. Towards the telomere, this block of sequences is followed by other subtelomeric sequences duplicated on several chromosomes (Fig. S3). Overlapping BACs 709I17 and 705I9 span from the FRG2C gene to subtelomeric sequences downstream to the clusters of D4Z4 and Beta satellite on chromosomes 4q35.2/10q26.3, where they are linked to approximately 110 kb of 3p12.3 sequences. The approximate length of D4Z4 clusters was derived from the estimated length of the analyzed BACs obtained by PFGE (not shown): all the sequenced BACs seem to carry a D4Z4 cluster of 15–20 repetitive units.

The organization of analyzed BACs is depicted in Fig. 2: BACs 584M21 and 809J1 span approximately 150 kb of the chimpanzee 4qter region, whereas approximately 200 kb of the chimpanzee 3p12

chromosome are present in BACs 709I17 and 705I9. Thus, in chimpanzee 4q35 and 3p12 chromosomes carry a very similar repetitive sequence organization.

3.5. Comparison of D4Z4 and Beta satellite repeats from PTR and HSA 4q35 chromosome

The employed sequencing approach did not allow to obtain full length D4Z4 units; this was probably due to difficulties in assembling the sequences of this tandem array, and/or to underrepresentation of D4Z4 sequences during the amplification step needed for the sequencing process. Nevertheless, by aligning several D4Z4 sequenced fragments from both 3p12 and 4q35 chromosome regions we derived two almost complete D4Z4 repeats labeled as D4Z4 4q (2992 bp) and D4Z4 3p (3263 bp) (accession numbers JN851147 and JN851148, respectively). These D4Z4 sequences (4q and 3p) contain an ORF of 1266 bp or 1272 bp, giving putative DUX4 protein of 422 (DUX4 4q) and 424 (DUX4 3p) aminoacids, respectively. Both protein sequences show a high percentage of identity (100% and 94.8% for DUX 4q and DUX 3p, respectively) with a previously reported putative DUX4 protein from Pan troglodytes (accession number CAL41938). Furthermore, PTR DUX 4q and 3p showed an identity of 95.8% and 92.5%, with a human DUX protein (accession number ADK24688). The alignment of chimpanzee and human DUX4 proteins is reported in Fig. 3.

The nucleotide consensus sequences of PTR and HSA Beta satellites were then derived from 21 and 77 repeat units present, respectively, in the cluster contained within human clone U774496 (4qA) and chimpanzee BAC 584M21(4q) (Fig. 4). The two consensus showed a high degree of identity (98.5%), as also derived from the frequency profiles of the most represented base at each position (Fig. 4). We then compared two other PTR and HSA sequences, namely pLAM and SSLP, which are considered important markers of the human 4q35 region [9,10]. pLAM sequences present in chimpanzee 709I17 and 584M21 BACs exhibited an identity of 68% and 93%, respectively, when compared with the human pLAM (accession number Z252821) (not shown); furthermore both PTR pLAMs do not contain any poly(A)
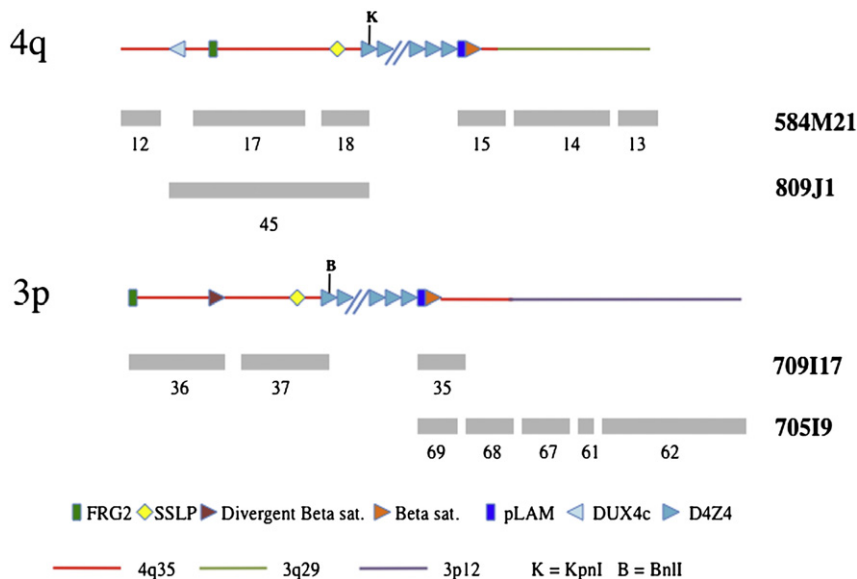

Fig. 2. Schematic sequence organization of PTR BACs 584M21, 809J1, 705I9 and 709I17. Gray rectangles and numbers beneath the schematic organization identify the sequenced subregions (see Supplementary Table 4). The internal organization of each BAC has been derived by sequence alignment to human genome databank (hg19), and by the reciprocal sequence identity as for the pairs of BACs 584M21/809J1, and 1705I9/709I17. Chromosome location at 4q and 3p of the sequenced BACs was derived by FISH analysis and by nucleotide sequence identity. In this regard, within BAC 584M21, 4q35-specific subtelomeric sequences (in red) are linked to PTR-specific subtelomeric sequences (light green), present in the human genome on several subtelomeres as 3q29, 1p36, 5q35, and 19p13, but not at 4q35. In addition, BAC 705I9, which overlaps with BAC 709I17, shows a linkage of 4q35 subtelomeric sequences with interstitial 3p12 sequences (in violet). // indicates interruption in the D4Z4 array that for these BACs has been estimated to be of approximately 15 copies of the repetitive unit.
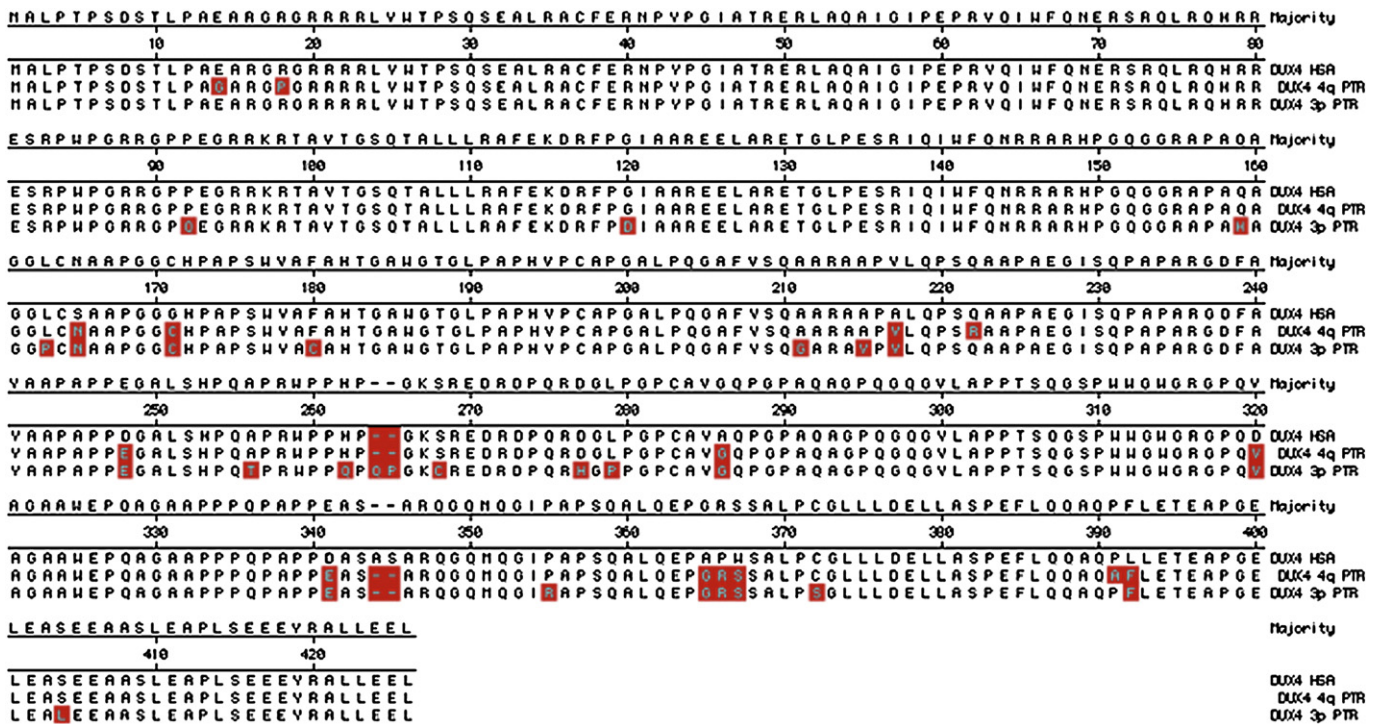
Fig. 3. Comparison of human and chimpanzee sequence features of D4Z4. Alignment of HSA and PTR 4q and 3p DUX4 amino-acid sequences. The three proteins show a length of 424 aa (HSA and 3p PTR) and 422 aa (4q PTR). The DUX 4q PTR protein showed a 100% identity with a previously reported chimpanzee DUX protein (accession number CAL41938).

signal (not shown). As regard the SSLP sequence located upstream to the D4Z4 array, 709I17 and 584 M21 BACs showed a nucleotide core of respectively 160 and 173 bases (not shown).

4. Discussion

The data here presented, not only provide a detailed analysis of the chimpanzee genome, but also add new insights into the evolutionary history of the human 4q35 chromosome region, and open the possibility that other repetitive sequences, in addition to the D4Z4 array, play a role in the functionality of this genomic region. The analysis of a PTR genomic library for clones highly positive for the D4Z4 repeat yielded a total of thirteen genomic sectors that were subsequently grouped into different classes: clones containing a linkage of 4q35 sequences with

sequences located in different human subtelomeres (i.e. 1p36.33 and 19p13.3), clones with a linkage of 4q35-like sequences and 3p12.3 sequences, and finally clones completely defined at both extremities by D4Z4 repeats. Furthermore, the PTR genomic clones exhibited in the D4Z4 unit a different restriction enzyme periodicity (KpnI or BlnI), and most importantly four of them, namely 584M21, 578I19, 709I17 and 705I9, carried a Beta satellite cluster linked to the D4Z4 array. Contrary to what is found in PTR and humans, in GGO this organization was not present on chromosome 4q but probably only on chromosome 4p. This conclusion was derived by the molecular analysis of the gorilla genome in which 4qter-derived genomic clones did not show the presence of Beta satellite sequences [25], and by FISH that showed the co-localization of D4Z4 and Beta satellite only on an internal location of chromosome 4p [28,25]. Furthermore, no Beta satellite
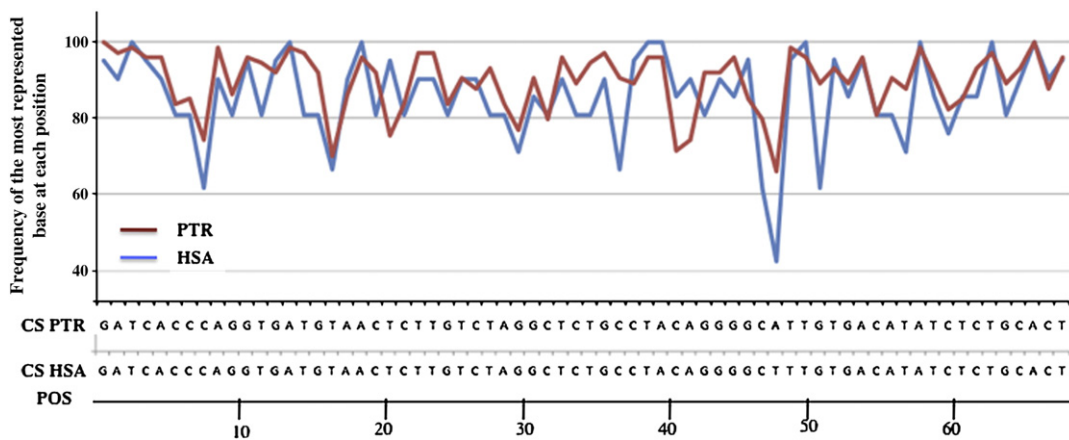

Fig. 4. Frequency of the most represented base at each position showed by Beta satellite DNA from PTR (red line) and HSA (blue line) as derived by the alignment of 74 repetitive units present in the cluster contained within PTR BAC 584M21-4q and of 21 repetitive units within the human clone U774496-4qA. Alignment of the consensus sequence of Beta satellite from PTR (CS PTR) and from HSA (CS HSA) derived from the values shown in the upper graphic. The alignment of two consensus showed a high degree of conservation (identity of 98.5%). POS = position of nucleotides within the repetitive unit.

sequences were detected by the screening of a rhesus monkey genomic library, and gibbon and orangutan Beta satellite sequences do not show a clustered organization [29]. Thus we suggest that, at 4qter, D4Z4 and Beta satellite linkage has recently appeared in the evolution, since this organization is only detected starting from the chimpanzee. This hypothesis is also supported by previous data reporting that the chimpanzee 4qter region is characterized by a sequence organization very similar to that found in humans [30]. Furthermore, in both studies only one type of 4qter organization was detected, and this was structurally similar to human 4qA up to a point just distal of the Beta satellite array, and included 4q specific subtelomeric sequences. Since this result was obtained by the screening of two independent PTR genomic libraries, it is reasonable to propose that PTR possesses only one type of 4qter organization belonging to the 4qA allele. A very similar sequence organization was detected on PTR chromosome 3p, where D4Z4 and Beta satellite linked arrays were embedded in 4q35-like sequences, spanning from the proximal FRG2C sequence to the distal subtelomeric sequences. Differences between the two chromosome regions included the occurrence on 3p of a divergent array of Beta satellite located upstream to the D4Z4 array, the restriction enzyme periodicity of D4Z4 (BlnI), and the number of PstI sites present in the Beta satellite linked to D4Z4. Starting from these findings in the PTR and from the organization of human syntenic regions, we derived a hypothetical evolutionary history of 4q35 and 3p12.2 that occurred after the divergence of chimpanzee from the human lineage (Fig. 5). An assumption is that a sequence organization composed of D4Z4 and Beta satellite linked arrays (4qA-like) was present at 4qter and 3p in a common ancestor of PTR and HSA. In HSA lineage, one of the two 4qA-like sequence organizations underwent the deletion of a region including one D4Z4, the pLAM and the Beta satellite cluster, originating the 4qB allele. Due to the identity of 97.8% between this PTR 4q region and human chromosome 18p11.32, it is possible to hypothesize that the deleted region has been transposed into the subtelomere of chromosome 18p. This hypothesis, in agreement with previous findings derived by the comparison of subtelomeric regions from human chromosomes 4qA and 4qB [11], suggests that the 4qA allele is evolutionary

older than 4qB. A duplication event in PTR chromosome 4q interesting a region spanning from the FRG2 gene to the subtelomeric sequences, and including D4Z4 and Beta satellite linked arrays, might have originated human chromosome 10q. Human chromosomes 3p12.3 and 4q35.2 show sequence identity in a region of approximately 35–40 kb comprised between one inverted D4Z4 (probably a copy of DUX4c) upstream to FRG2, and sequences just proximal to the D4Z4 array. Hence, human chromosome 3p12 might have originated by the total deletion of D4Z4 and Beta satellite linked arrays and of 4q35 subtelomeric sequences (Fig. 5). Alternatively, D4Z4 and Beta satellite linked arrays were not present at 3p in a common ancestor of PTR and HSA, and thus this repetitive sequence organization appeared in the PTR lineage. Thus, after human/chimpanzee divergence D4Z4 and Beta satellite linked arrays underwent an extensive genomic remodeling leaving this sequence organization at the subtelomeric location of human chromosomes 4qA and 10q26. In this regard, search in human reference sequence (hg19) and in human unplaced contigs and clones for D4Z4 and Beta satellite positivity, strongly indicates that the linking of the two families of repeats is a peculiar feature of chromosomes 4q and 10q. In fact, as previously reported [31] particularly for the p arm of acrocentric chromosomes and for the Y chromosome, arrays of Beta satellite are intermingled with single or double D4Z4 repeats.

A different evolutionary history characterizes the two families of repeats and their reciprocal sequence organization typical of human chromosomes 4q and 10q. D4Z4 repeats were not confined to primates, but found in several placental mammals thus indicating an ancient origin [32]. All the analyzed species showed the maintenance of the DUX ORF and an array organization containing at least 10 repeated units. Furthermore a survey of non-mammalian genomes identified a likely ancestor of the mammalian DUX gene [33]. Conversely, the appearance of clustered Beta satellite sequences can be traced back to the gorilla, where the bulk of the repeats were mapped interstitially to 4p and to the p arm of chromosome Y [28,29]. That the gorilla 4qter region contains clustered D4Z4 units but not Beta satellite linked array was confirmed by genome analysis [25]. Hence, D4Z4 and Beta satellite linked
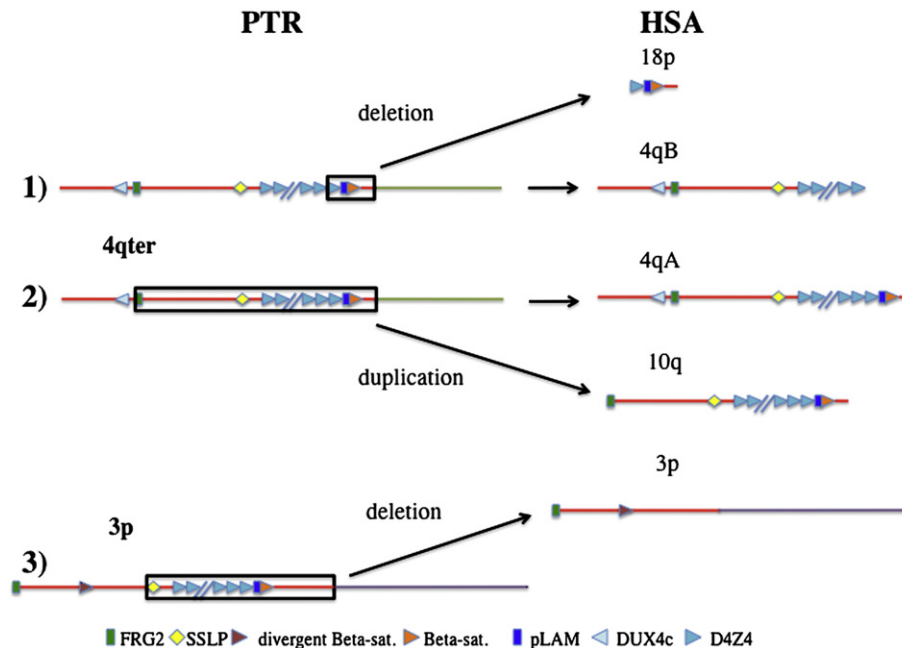

Fig. 5. Schematic representation of the hypothetical evolutionary events occurring in the human lineage after the separation of chimpanzee. The PTR 4qter chromosome region is represented with two copies of the 4qA-like allele, whereas the interstitial 3p region with only one copy. (1) One 4qA-like allele underwent to an event of deletion and transposition to chromosome 18pter of a region comprising one D4Z4, pLAM, the Beta satellite array and subtelomeric sequences (boxed region on the upper 4qA-like copy). This deletion originated the human 4qB allele, whereas the other 4qter allele became allele 4qA. (2) One 4qA-like allele was used as template for the duplication of a region spanning from FRG2 to subtelomeric sequences, and including the full D4Z4 and Beta satellite linked arrays; this duplication was translocated to the qter region of chromosome 10. (3) The interstitial D4Z4 and Beta satellite linked arrays at PTR 3p12 was deleted (boxed region) on both chromosomes leaving upstream sequences just to the DUX4c and including the FRG2C gene and a short array of divergent Beta satellite.

arrays at 4qter represents a recent evolutionary event occurring in the PTR after the divergence of the gorilla lineage. However, during human evolution these repetitive regions underwent further genomic remodeling that caused the loss of the Beta satellite cluster from one 4q allele and the appearance of D4Z4–Beta satellite linked organization at 10q. Nevertheless, as in PTR 4q and 3p, human D4Z4 units maintain a coding capacity at 4q as well as at 10q. The importance of this repetitive sequence organization for human cell physiology can be derived from the observation that D4Z4 array contraction of the 4qA allele causes a type of muscle dystrophy (FSHD) [8–10]. There is a general consensus in the field that D4Z4 deletion leads to epigenetic alterations that affect the expression profiles of in cis candidate genes (for review see [9,10,34]). Potential FSHD candidate genes are ANT1 (SLC25A4), FRG1, FRG2, DUX4 and DUX4c and in this regard in the last years many papers have attempted to define the "true" candidate gene(s) without conclusive results. Although DUX4 and FRG1 seem to be the main candidate genes, the heterogeneity of disease manifestations could probably underline a heterogeneity of gene expression due to different epigenetic alterations of in cis candidate genes. It has been previously demonstrated that both contracted and non contracted 4q35 alleles associate with the heterochromatic nuclear compartment (nuclear rim and nucleolar periphery) [35,36]. These studies also indicated that other chromosomal loci bearing the D4Z4 repeats, such as 10q and the p arm of acrocentric chromosomes showed a common affinity for these heterochromatic regions. Furthermore, a common feature of these D4Z4 chromosome regions is that they reside adjacent to blocks of Beta satellite DNA. Indeed, D4Z4 does not seem to be responsible for the localization to the heterochromatic compartment, but rather for the transcriptional repression of the region via heterochromatinization. In this regard, it has been demonstrated that in normal cells D4Z4 repeats are methylated and show histone marks of repression (H3K27me3), and that both features are lost in FSHD patients [16,17,37]. Furthermore a repressor complex composed of YY1, HMGB2 and nucleolin is specifically bound to D4Z4 [38,39]. For this reason a possible role of Beta satellite in augmenting the observed heterochromatic repressive signal at the 4q35 region by acting synergistically with the D4Z4 array in recruiting chromatin modifiers can be hypothesized. At present no data are available either on the polymorphism of 4q35 Beta satellite array within the human population, or whether the array might undergo contraction like the D4Z4 array. However, differently from the 4q35 D4Z4 array showing in the human population a great length polymorphism (ranging from 11 to 150 copies) [40,41], the few available information from the databank suggest a conservation in PTR and HSA of the length of 4q35 Beta satellite array, accounting for approximately 90 copies (6 kb). Moreover, we found that some subregions of the Beta satellite unit seem to be highly conserved between PTR and HSA. Although further computational and functional experiments are certainly required to confirm the conservation of Beta satellite sequence, our results suggest a possible biological role for this class of repeats. Interestingly, the obtained results are very similar to those previously reported by us [42]. In this regard, the observed variability of D4Z4-contracted FSHD manifestation, as well as FSHD manifestation in non-contracted individuals (phenotypic FSHD) and the absence of disease manifestation in carrier of D4Z4 contraction might be explained in term of Beta satellite DNA. In conclusion, these results have delineated the evolutionary history of linked D4Z4 and Beta satellite arrays, raising the possibility that Beta satellite repeats might be involved in the regulation of the FSHD region.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2012.07.011.

Acknowledgments

References

[1] W.F. Doolittle, C. Sapienza, Selfish genes, the phenotype paradigm and genome evolution, Nature 284 (1980) 601–603.
[2] L.E. Orgel, F.H. Crick, Selfish DNA: the ultimate parasite, Nature 284 (1980) 604–607.
[3] N.V. Tomilin, Regulation of mammalian gene expression by retroelements and non-coding tandem repeats, Bioessays 30 (2008) 338–348.
[4] D. Ugrakovic, Functional elements residing within satellite DNAs, EMBO Rep. 6 (2005) 1035–1039.
[5] C. Guetg, P. Lienemann, V. Sirri, I. Grummt, D. Hernandez-Verdun, M.O. Hottiger, M. Fussenegger, R. Santoro, The NoRC complex mediates the heterochromatin formation and stability of silent rRNA genes and centromeric repeats, EMBO J. 29 (2010) 2135–2146.
[6] G. Iotti, E. Longobardi, S. Masella, L. Dardaei, F. De Santis, N. Micali, F. Blasi, Homeodomain transcription factor and tumor suppressor Prep1 is required to maintain genomic stability, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) E314–E322.
[7] B. Mravinac, M. Plohl, D. Ugarkovic, Preservation and high sequence conservation of satellite DNAs suggest functional constraints, J. Mol. Evol. 61 (2005) 542–550.
[8] J.M. Statland, R. Tawil, Facioscapulohumeral muscular dystrophy: molecular pathological advances and future directions, Curr. Opin. Neurol. 24 (2011) 423–428.
[9] S.M. van der Maarel, R. Tawil, S.J. Tapscott, Facioscapulohumeral muscular dystrophy and DUX4: breaking the silence, Trends Mol. Med. 17 (2011) 252–258.
[10] D.S. Cabianca, D. Gabellini, The cell biology of disease: FSHD: copy number variations on the theme of muscular dystrophy, J. Cell Biol. 191 (2010) 1049–1060.
[11] M. van Geel, M.C. Dickson, A.F. Beck, D.J. Bolland, R.R. Frants, S.M. van der Maarel, P.J. de Jong, J.E. Hewitt, Genomic analysis of human chromosome 10q and 4q telomeres suggests a common origin, Genomics 79 (2002) 210–217.
[12] B.T. Buzhov, R.J. Lemmers, I. Tournev, M.J. van der Wielen, B. Ishpekova, R. Petkov, J. Petrova, R.R. Frants, G.W. Padberg, S.M. van der Maarel, Recurrent somatic mosaicism for D4Z4 contractions in a family with facioscapulohumeral muscular dystrophy, Neuromuscul. Disord. 15 (2005) 471–475.
[13] R.J. Lemmers, P. de Kievit, L. Sandkuijl, G.W. Padberg, G.J. van Ommen, R.R. Frants, S.M. van der Maarel, Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere, Nat. Genet. 32 (2002) 235–236.
[14] S.M. van der Maarel, R.R. Frants, G.W. Padberg, Facioscapulohumeral muscular dystrophy, Biochim. Biophys. Acta 1772 (2007) 186–194.
[15] G. Jiang, F. Yang, P.G. van Overveld, V. Vedanarayanan, S. van der Maarel, M. Ehrlich, Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q, Hum. Mol. Genet. 12 (2003) 2909–2921.
[16] P.G. van Overveld, R.J. Lemmers, L.A. Sandkuijl, L. Enthoven, S.T. Winokur, F. Bakels, G.W. Padberg, G.J. van Ommen, R.R. Frants, S.M. van der Maarel, Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy, Nat. Genet. 35 (2003) 315–317.
[17] J.C. de Greef, R.J. Lemmers, B.G. van Engelen, S. Sacconi, S.L. Venance, R.R. Frants, R. Tawil, S.M. van der Maarel, Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD, Hum. Mutat. 3 (2009) 1449–1459.
[18] J.C. de Greef, R.J. Lemmers, P. Camano, J.W. Day, S. Sacconi, M. Dunand, B.G. van Engelen, S. Kiuru-Enari, G.W. Padberg, A.L. Rosa, C. Desnuelle, S. Spuler, M. Tarnopolsky, S.L. Venance, R.R. Frants, S.M. van der Maarel, R. Tawil, Clinical features of facioscapulohumeral muscular dystrophy 2, Neurology 75 (2010) 1548–1554.
[19] E. Ballestar, A.P. Wolffe, Methyl-CpG-binding proteins. Targeting specific gene repression, Eur. J. Biochem. 268 (2001) 1–6.
[20] J.C. de Greef, M. Wohlgemuth, O.A. Chan, K.B. Hansson, D. Smeets, R.R. Frants, C.M. Weemaes, G.W. Padberg, S.M. van der Maarel, Hypomethylation is restricted to the D4Z4 repeat array in phenotypic FSHD, Neurology 69 (2007) 1018–1026.
[21] S. Pandya, W. King, R. Tawil, Facioscapulohumeral dystrophy, Phys. Ther. 88 (2008) 105–113.
[22] R. Tawil, S.M. van der Maarel, Facioscapulohumeral dystrophy, Muscle Nerve 34 (2008) 1–15.
[23] K.L. Bastress, J.M. Stajich, M.C. Speer, J.R. Gilbert, The genes encoding for D4Z4-binding proteins HMGB2, YY1, NCL and MYOD are excluded as candidate genes for FSDH1B, Neuromuscul. Disord. 15 (2005) 316–320.
[24] M.M. Tonini, M.R. Passos-Bueno, A. Cerqueira, S.R. Matioli, R. Pavanello, M. Zatz, Asymptomatic carriers and gender differences in facioscapulohumeral muscular dystrophy (FSHD), Neuromuscul. Disord. 14 (2004) 33–38.
[25] B. Bodega, M.F. Cardone, S. Muller, M. Neusser, F. Orzan, E. Rossi, E. Battaglioli, A. Marozzi, P. Riva, M. Rocchi, R. Meneveri, E. Ginelli, Evolutionary genomic remodelling of the human 4q subtelomere (4q35.2), BMC Evol. Biol. 7 (2007) 39–51.
[26] R. Meneveri, A. Agresti, G. Della Valle, D. Talarico, A.G. Siccardi, E. Ginelli, Identification of a human clustered G+C-rich DNA family of repeats (Sau3A family), J. Mol. Biol. 186 (1985) 483–489.
[27] Primate Cytogenetics, http://www.biologia.uniba.it/primates/cell_lines.html.
[28] H. Hirai, T. Taguchi, A.K. Godwin, Genomic differentiation of 18S ribosomal DNA and Beta satellite DNA in the hominoid and its evolutionary aspects, Chromosome Res. 7 (1999) 531–540.
[29] M.F. Cardone, L. Ballarati, M. Ventura, M. Rocchi, A. Marozzi, E. Ginelli, R. Meneveri, Evolution of beta satellite DNA sequences: evidence for duplication-mediated repeat amplification and spreading, Mol. Biol. Evol. 21 (2004) 1792–1799.
[30] M.K. Rudd, R.M. Endicott, C. Friedman, M. Walker, J.M. Young, K. Osoegawa, NISC Comparative Sequencing Program, P.J. de Jong, E.D. Green, B.J. Trask, Comparative sequence analysis of primate subtelomeres originating from a chromosome fission event, Genome Res. 19 (2009) 33–41.
[31] S.T. Winokur, U. Bengtsson, J.C. Vargas, J.J. Wasmuth, M.R. Altherr, B. Weiffenbach, S.J. Jacobsen, The evolutionary distribution and structural organization of the

homeobox-containing repeat D4Z4 indicates a functional role for the ancestral copy in the FSHD region, Hum. Mol. Genet. 5 (1996) 1567–1755.

[32] J. Clapp, L.M. Mitchell, D.J. Bolland, J. Fantes, A.E. Corcoran, P.J. Scotting, J.A. Armour, J.E. Hewitt, Evolutionary conservation of a coding function for D4Z4, the tandem DNA repeat mutated in facioscapulohumeral muscular dystrophy, Am. J. Hum. Genet. 81 (2007) 264–279.

[33] A. Leidenroth, J.E. Hewitt, A family history of DUX4: phylogenetic analysis of DUXA, B, C and Duxbl reveals the ancestral DUX gene, BMC Evol. Biol. 10 (2010) 364–376.

[34] M.V. Neguembor, D. Gabellini, In junk we trust: repetitive DNA, epigenetics and facioscapulohumeral muscular dystrophy, Epigenomics 2 (2010) 271–287.

[35] R. Tam, K.P. Smith, J.B. Lawrence, The 4q subtelomere harboring the FSHD locus is specifically anchored with peripheral heterochromatin unlike most human telomeres, J. Cell Biol. 167 (2004) 269–279.

[36] P.S. Masny, U. Bengtsson, S.A. Chung, J.H. Martin, B. van Engelen, S.M. van der Maarel, S.T. Winokur, Localization of 4q35.2 to the nuclear periphery: is FSHD a nuclear envelope disease? Hum. Mol. Genet. 13 (2004) 1857–1871.

[37] W. Zeng, J.C. de Greef, Y.Y. Chen, R. Chien, X. Kong, H.C. Gregson, S.T. Winokur, A. Pyle, K.D. Robertson, J.A. Schmiesing, V.E. Kimonis, J. Balog, R.R. Frants, A.R. Jr Ball, L.F. Lock, P.J. Donovan, S.M. van der Maarel, K. Yokomori, Specific loss of histone

H3 lysine 9 trimethylation and HP1gamma/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral muscular dystrophy (FSHD), PLoS Genet. 5 (2009) e1000559.

[38] D. Gabellini, M.R. Green, R. Tupler, Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle, Cell 110 (2002) 339–348.

[39] B. Bodega, G.D. Ramirez, F. Grasser, S. Cheli, S. Brunelli, M. Mora, R. Meneveri, A. Marozzi, S. Mueller, E. Battaglioli, E. Ginelli, Remodeling of the chromatin structure of the facioscapulohumeral muscular dystrophy (FSHD) locus and upregulation of FSHD-related gene 1 (FRG1) expression during human myogenic differentiation, BMC Biol. 7 (2009) 41.

[40] C. Wijmenga, J.E. Hewitt, L.A. Sandkuijl, L.N. Clark, T.J. Wright, H.G. Dauwerse, A.M. Gruter, M.H. Hofker, P. Moerer, R. Williamson, G.J. van Ommen, G.W. Padberg, R.R. Frants, Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy, Nat. Genet. 2 (1992) 26–30.

[41] J.C. van Deutekom, C. Wijmenga, E.A. van Tienhoven, A.M. Gruter, J.E. Hewitt, G.W. Padberg, G.J. van Ommen, M.H. Hofker, R.R. Frants, FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit, Hum. Mol. Genet. 2 (1993) 2037–2042.

[42] A. Agresti, R. Meneveri, A.G. Siccardi, A. Marozzi, G. Corneo, S. Gaudi, E. Ginelli, Linkage in human heterochromatin between highly divergent Sau3A repeats and a new family of repeated DNA sequences (HaeIII family), J. Mol. Biol. 205 (1989) 625–631.