The American Journal of Human Genetics Volume 86

Supplemental Data

Worldwide Population Analysis of the 4q and 10q Subtelomeres Identifies Only Four Discrete Interchromosomal Sequence Transfers in Human Evolution

Richard J.L.F. Lemmers, Patrick J. van der Vliet, Kristiaan J. van der Gaag, Sofia Zuniga, Rune R. Frants, Peter de Knijff, and Silvère M. van der Maarel

Figure S1

Repeat array composition of typical 4A166, 4A166H and 4A-Hi chromosomes.

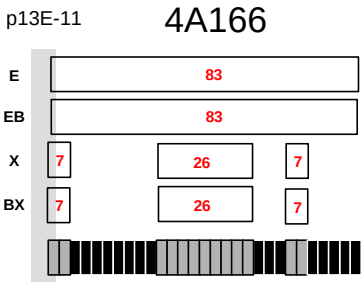
The composition of the D4Z4 repeat array in four different hybrid haplotypes as inferred from Southern blotting of restriction enzyme digested genomic DNA. For each hybrid haplotype we selected individuals carrying one hybrid chromosome (boxed in Southern blot) and three standard chromosomes (one chromosome 4 and two chromosomes 10). The Southern blot was hybridized with probe p13E-11 (recognizing the region proximal to D4Z4, see figure 2) and probe D4Z4 (recognizing individual D4Z4 units). The first digestion (E, double digestion of *Eco*RI and *Hin*dIII) shows all four D4Z4 repeat arrays from chromosomes 4 and 10. The EB lane (double digestion with *Bln*I and *Eco*RI) reveals all repeat arrays that are resistant to *Bln*I (consisting of B^-X^- and B^-X^+ units). Similarly, the third lane (X) shows *Xap*I resistant arrays (consisting of B^-X^- and B^+X^- units). The *Xap*I and *Bln*I double digestion (BX) in the fourth lane identifies B^-X^- repeat arrays only. On the left side of the p13E-11 hybridized Southern blot the array sizes of the standard (black) and hybrid repeat arrays (red) are indicated. On the right side of the D4Z4 hybridized Southern blot the residual array fragments of the EB-, X- and BX-digested hybrid repeat array are indicated in red. The left panel summarizes the Southern results with the different fragment sizes for each digestion (E, EB, X and BX), it shows one of the possible compositions of the hybrid D4Z4 repeat array and the sizes of the three standard repeat arrays from 4q and 10q.

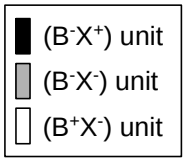
The sequences of the SSLP and D4F104S1 region proximal to the repeat (figure 4) are identical for the chromosomes shown in a, b and c and also identical to the standard chromosome 10A166. The sequence of the 4A161Hi chromosome is identical to standard 4A161 chromosomes in the SSLP and D4F104S1 regions (figure 4). The difference between these different hybrid repeat arrays is most evident in the boxed segment and the fragment indicated by an asterisk in the Southern blot. The boxed segment indicates the hybrid repeat at

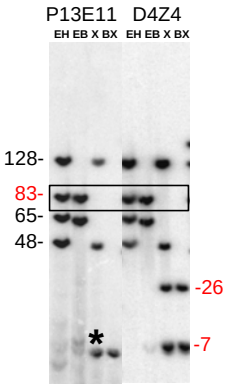
different digestions and the asterisk on the Southern blot probed with p13E-11 marks the sequence features of the D4Z4 unit for the most proximal D4Z4 units of the array (figure a). The 4A166 array is completely resistant to *Bln*I (B^-) (boxed) and the two most proximal units are in addition also resistant to *Xap*I (asterisk). Additional B^-X^- arrays are visible upon D4Z4 hybridization in the X lane. Thus 4A166 arrays consist of arrays of B^-X^- and of B^-X^+ units. The 4A166H(a) D4Z4 array in figure b carries two B^+X^- units at the most proximal position (asterisk in X lane) and the remainder of the array is completely resistant to *Bln*I (boxed region on D4Z4 blot in B lane). This 4A166H D4Z4 array has a similar composition as 4A166, but with some B^+X^- units in most proximal region. The 4A166H(b,c) array in figure c also carries two B^+X^- units at the most proximal position (asterisk in X lane), but the remainder of the array is in contrast to 4A166H(a) interrupted by other B^+X^- units. As a consequence, no residual fragment is visible in the box upon D4Z4 hybridization in lane B, but instead three smaller *Bln*I resistant fragments (51, 15 and 10 kb) are visible. Thus in summary, 4A166H(b,c) D4Z4 arrays have a similar composition as in 4A166H(a), but also carries internal B^+X^- arrays.

The array of 4A161Hi chromosomes (figure d) differs completely from the three previous arrays. The most proximal D4Z4 units are B^-X^+ for the first 55 kb (asterisk in EB lane) similar to standard 4A161 chromosomes (like the region proximal of 4A161Hi chromosomes). The repeat array consist of three B^-X^+ arrays (105, 55 and 9 kb) as shown in lane EB upon D4Z4 hybridization and is interrupted by some B^+X^- units.

a



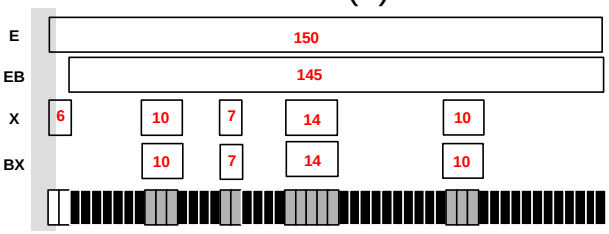


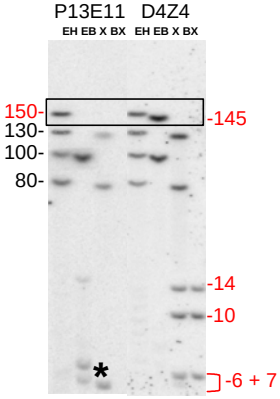


Standard 4q: 78 kb

Standard 10q: 63 kb and 160 kb

b

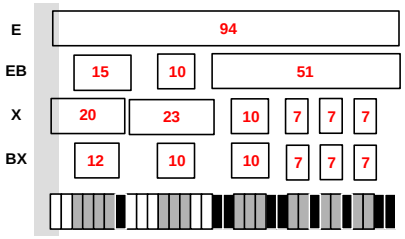


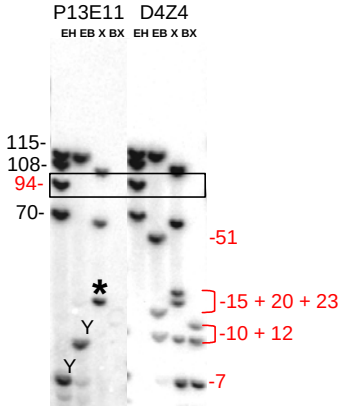


Standard 4q: 100 kb

Standard 10q: 80 kb and 130 kb

c

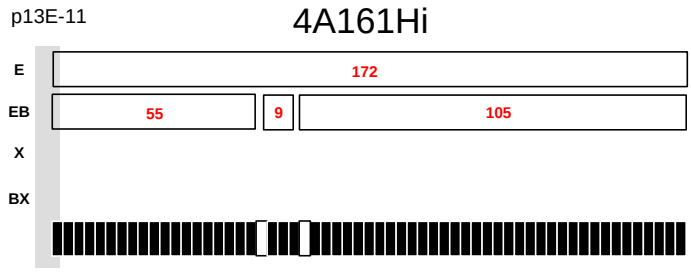


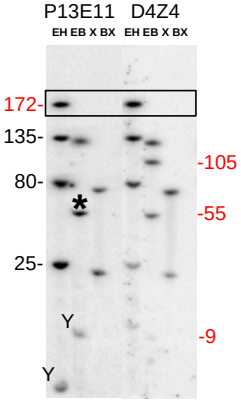


Standard 4q: 115 kb

Standard 10q: 70 kb and 108 kb

d





Standard 4q: 135 kb

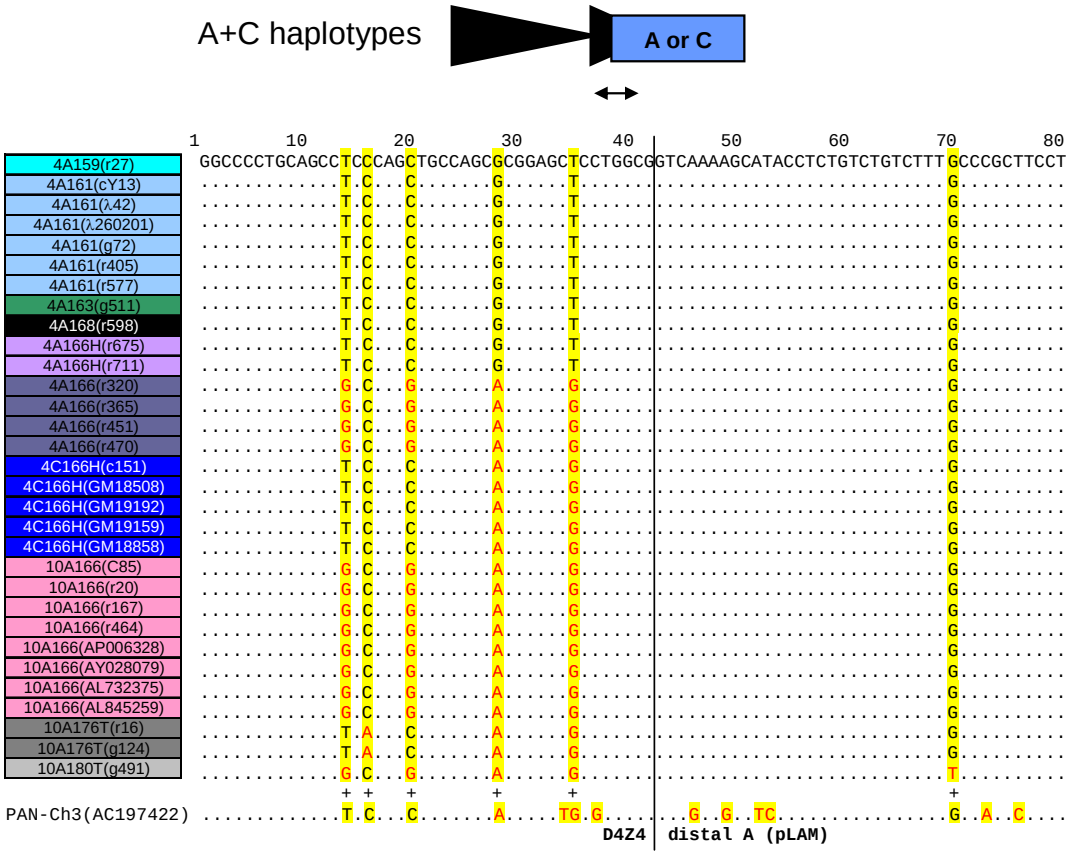
Standard 10q: 25 kb and 80 kb

Figure S2

Breakpoint between D4Z4 and distal region on A+C and B chromosomes.

a) Breakpoint between D4Z4 and the region immediately distal to D4Z4 on A and C chromosomes of different key haplotypes. Chromosomes have been sequenced that belong to six different 4A haplotypes (4A159, 4A161, 4A163, 4A168, 4A166 and 4A166H), three 10A haplotypes (10A166, 10A176T and 10A180T) and to the 4C166H haplotype. The sample number (LUMC identifier or Coriell catalog ID) from which the chromosome was isolated is shown between brackets. Some 10A166 sequences and the chimpanzee sequence were obtained from Genbank and the accession numbers are between brackets. Sequence variations compared to the most common 4A161 haplotype are shown in red.

b) Breakpoint between D4Z4 and region immediately distal to D4Z4 on B chromosomes of different key haplotypes (4B163, 4B168 and 10B161T). The LUMC sample number from which the chromosome was isolated is shown between brackets. The sequence from accession number AF017466 represents the 4B163 sequence from Genbank. No sequence variations have been identified between chromosomes of these specific haplotypes.


A+C haplotypes
A or C
1 10 20 30 40 50 60 70 80
4A159(r27) GGCCCCTGCAGCCTCCCAGCTGCCAGCGCGGAGCTCCTGGCGGTCAAAAGCATACCTCTGTCTGTCTTTGCCCGCTTCCT
4A161(cY13)
4A161(λ42)
4A161(λ260201)
4A161(g72)
4A161(r405)
4A161(r577)
4A163(g511)
4A168(r598)
4A166H(r675)
4A166H(r711)
4A166(r320)
4A166(r365)
4A166(r451)
4A166(r470)
4C166H(c151)
4C166H(GM18508)
4C166H(GM19192)
4C166H(GM19159)
4C166H(GM18858)
10A166(C85)
10A166(r20)
10A166(r167)
10A166(r464)
10A166(AP006328)
10A166(AY028079)
10A166(AL732375)
10A166(AL845259)
10A176T(r16)
10A176T(g124)
10A180T(g491)
PAN-Ch3(AC197422)
D4Z4
distal A (pLAM)

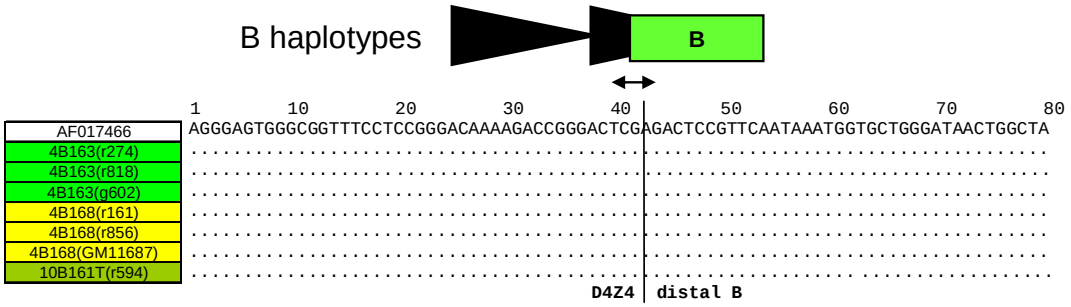

B haplotypes
B
1 10 20 30 40 50 60 70 80
AF017466 AGGGAGTGGGCGGTTTCCTCCGGGACAAAAGACCGGGACTCGAGACTCCGTTCAATAAATGGTGCTGGGATAACTGGCTA
4B163(r274)
4B163(r818)
4B163(g602)
4B168(r161)
4B168(r856)
4B168(GM11687)
10B161T(r594)
D4Z4
distal B

Figure S3

Neighbour joining tree of all 4q and 10 q sequence haplotypes.

An unrooted neighbour joining tree based on all distinct 4q and 10q haplotypes. We included the orthologous D4Z4 region present on chimpanzee chromosome 3 allowing us to infer the root of this tree. The indicated branch support values reflect the percentage of 1000 bootstrapped trees showing a similar branch. The topology of this tree is extremely similar to the median joining network (see figure 6) which is repeated here as a smaller insert.

The only difference is the branch with a rather low support of 49% which (not surprisingly) is positioned exactly at the small cluster of four haplotypes in the median joining network connected by a reticulating cube. This also illustrates the power of using networks instead of bifurcating trees: In a network all possible connections are shown without any assumption, whereas in a bifurcating tree such reticulating clusters (most often due to recurrent mutations) invariably lead to very low branch supports.

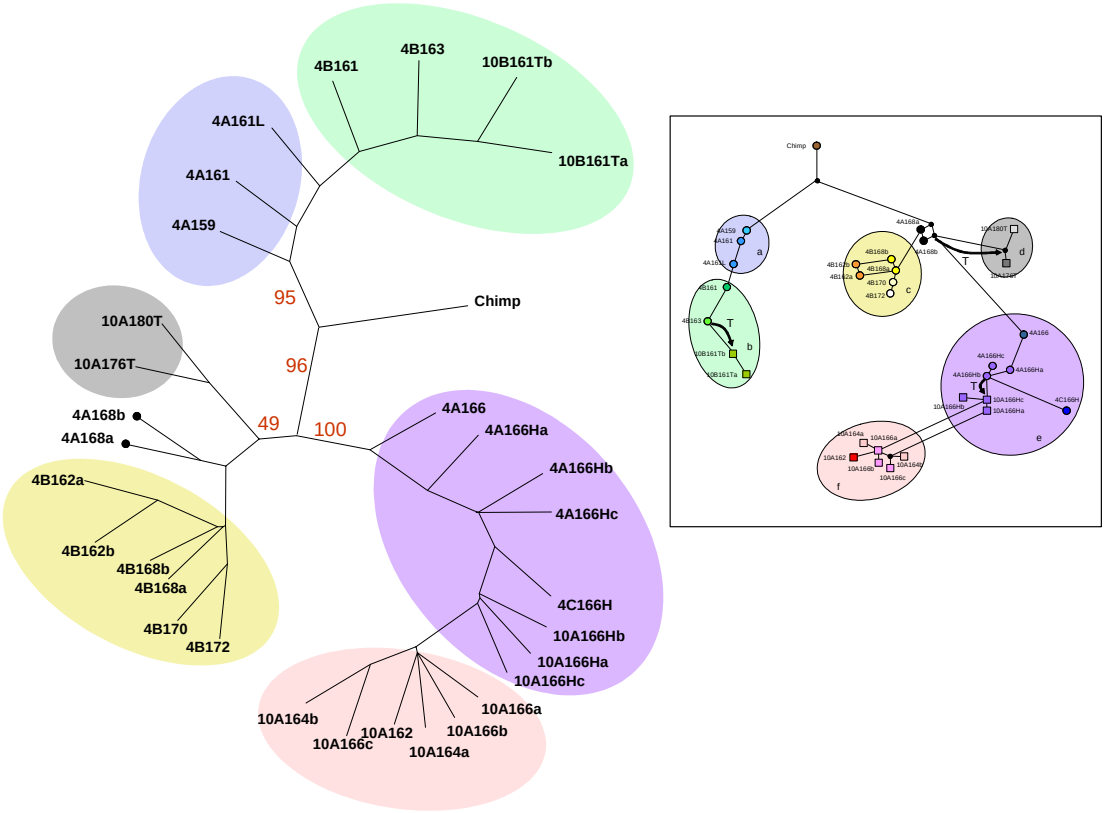


Figure S4

Correlation between genetic distance and genetic variation.

The genetic variation of the SSLP, expressed in term of average number of distinct alleles per locus per population (top two panels) and unbiased expected heterozygosity per locus per population (bottom two panels) are compared with the population specific geographic distance (expressed as the shortest Great-Circle distance via way-points) from East-Africa. For both loci there is a marked decrease in average number of alleles and for the 4q SSLP there is also a marked decrease in heterozygosity.