

Survey of maximum CTG/CAG repeat lengths in humans and non-human primates: total genome scan in populations using the Repeat Expansion Detection method

Giorgio Sirugo¹, Amos S. Deinard^{1,2}, Judith R. Kidd¹ and Kenneth K. Kidd^{1,*}

Yale University, ¹School of Medicine, Department of Genetics, and ²Department of Anthropology, New Haven, 06510, Connecticut, USA

Received September 17, 1996; Revised and Accepted December 31, 1996

Repeat Expansion Detection (RED) is an efficient and simple method for detecting repeat expansions in the human genome, including expansion mutations resulting in disease. Here we report the first population survey of CTG/CAG repeat lengths in humans using the RED method; we have determined maximum CTG/CAG repeat length in 244 individuals from six human populations: Danes, Chinese, Japanese, Rondonian Surui, Maya and Mbuti/Biaka Pygmies. We have also sampled a number of non-human primates including eight orang-utans (*Pongo pygmaeus*), seven gorillas (*Gorilla gorilla*), seven pygmy chimpanzees (*Pan paniscus*), 13 common chimpanzees (*Pan troglodytes*) and three *Hylobatidae* (one *Hylobates lar*, one *H.klossii*, and one *H.syndactylus*). Our results demonstrate the existence of significant variation in the sizes and frequencies of the longest CTG/CAG repeat length seen per individual both within and between human populations. The population differences argue that overall mutation rates at CTG/CAG repeat loci are sufficiently low that mutation does not obliterate the effect of random genetic drift and clearly indicate that population stratification could occur in disease association studies using the RED method. No significant differences were detected among the non-human primates sampled. Our results also show that both common chimpanzees and pygmy chimpanzees (bonobos) are polymorphic for maximum length of any CTG/CAG repeats while no variation was found for gorillas and orang-utans.

INTRODUCTION

Expansions of unstable DNA repeats have been demonstrated to be involved in ten different inherited disorders including fragile X syndrome, myotonic dystrophy and Huntington disease (1,2). Most of these trinucleotide repeat expansion disorders share

unusual genetic features: increasing penetrance and disease severity in successive generations (genetic anticipation) along with a parental (maternal or paternal) sex bias in the transmission of the severe form of the disease which correlates with the degree of meiotic instability and allelic expansion. The rather frequent occurrence of triplet repeats in mRNA indicates that more loci containing unstable DNA expansions could be discovered (1). Allowing for the complementary nature of DNA and frame permutations, ten possible triplet repeats can occur at the DNA level; thus far (CAG)_n (identical to (CTG)_n) repeats are involved in seven of the disorders already identified while the remaining disorders involve two with (CGG)_n and one with (GAA)_n repeats (2,3). At all of these loci polymorphic repeat lengths (means of ~20 repeats) correspond to the normal alleles while abnormal, disease-causing alleles are larger. When the repeats are located in coding regions of genes (six loci, all responsible for neurodegenerative syndromes), the disease alleles are in the 35–100 repeats range (2,3), while pathogenic expansions for the remaining disorders may be as 'small' as 50 repeats (4) and can extend to hundreds or thousands of repeats.

Recently, Schalling *et al.* (5) described a generally applicable method, termed Repeat Expansion Detection (RED), for detecting the presence of large, potentially unstable trinucleotide repeats. The method uses genomic DNA as a template for the annealing and ligation of repeat-specific oligonucleotides and does not require flanking sequence determination or single copy probes. The genomic DNA serves as a 'guide' and the maximum number of oligos that can be ligated is the number that can simultaneously anneal to the longest sequence, anywhere in the genome, with a homogenous repeat complementary to the specific trinucleotide(s) oligos used in the ligation reaction. The amount of ligation product generated is enhanced by using multiple cycles of annealing, ligation and denaturation. The size of the largest ligation product is the result recorded for each individual. The RED method has recently been used to explore the possibility that expansion of the number of repeats at CTG/CAG trinucleotide repeat loci could underlie neuropsychiatric disorders, especially those that might show genetic anticipation (6–8). To date, no information has been generated regarding the frequency distribution of long CTG/CAG trinucleotide repeats in humans

*To whom correspondence should be addressed

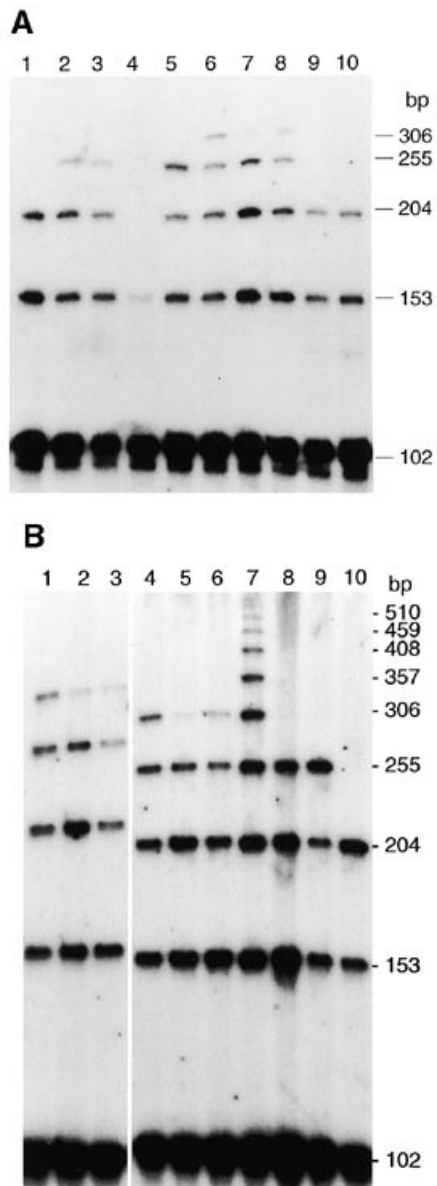


Figure 1. (A) RED of $(CTG)_{17}$ ligations in a sample of unrelated Chinese. The size of ligated multimers is indicated in bp. Note lanes 2, 3 and 5 with fragments up to 255 bp corresponding to the ligation of five $(CTG)_{17}$ oligos, lanes 6, 7, and 8 with six ligated $(CTG)_{17}$ (306 bp) and lane 4 with 153 bp corresponding to the ligation of only three oligonucleotides. (B) Lanes 1, 2, 3, and 4, 5, 6 show 2nd and 3rd replication of results in samples of unrelated Chinese (JK 3034, JK 3035, JK 3036) where $(CTG)_{102}$ were detected by RED [lanes 6, 7, 8 in (A)]. Both the maximum repeat length and the intensity of RED multimers are fully reproducible when the same templates are used in separate reactions, demonstrating that ligations of $(CTG)_{17}$ do not occur stochastically. Replicate RED analyses of four additional samples (lanes 7, 8, 9, 10) confirmed the maximum repeat length detected for the same samples in a previous set of ligations (not shown). The long CTG/CAG ladder in lane 7 was detected in a Japanese individual. Although particularly strong reactions were observed for all samples, no additional artefactual fragments were detected and identical sizes were observed in many replicate analyses of which these are illustrative.

other than northern Europeans (5,7,8). We have used the RED method for an extensive, total genome/genomic survey of CTG/CAG repeats in six human populations and a sample of non-human primates.

RESULTS

Using 5 μ g of genomic DNA for RED gives highly reproducible results as illustrated for the three samples represented in Figure 1A and B. The number of ligation products is also clearly a function of the genomic DNA sample used since the reactions for the samples in Figure 1A were run using aliquots of the same reaction cocktail and simultaneously cycled. The separate set of RED reactions illustrated in lanes 4–10 in Figure 1B were also identical, except for the genomic DNA.

Results of our survey are shown in Figures 2 and 3. The distributions of the maximum number of ligated $(CTG)_{17}$ oligos in each sample for different human populations and for non-human primates are shown as the number of $(CTG)_{17}$ ligation events on the X axis, and the relative frequency (human populations, Fig. 2) or the absolute frequency (non-human primates, Fig. 3) on the Y axis. We found that the distributions of largest CTG/CAG repeat lengths differed significantly among human populations ($\chi^2 = 107$, $p < 0.001$, 15 d.f.), while the differences detected among non-human primate species were not significant (2-tailed Fisher exact test, $p = 0.455$). Differences of frequency distributions of $(CTG)_n$ repeats in human populations were also tested by pairwise, non-parametric Kolmogorov-Smirnov analysis. Differences were highly significant ($p < 0.005$) for eight of the 15 pairwise comparisons analyzed, but the other seven were not significant: Chinese vs. Japanese ($p = 0.141$), Chinese vs. Maya ($p = 0.141$), Japanese vs. Surui ($p = 0.239$), Japanese vs. Maya ($p = 0.25$), Japanese vs. Danes ($p = 0.021$) and Surui vs. Maya ($p = 0.89$) pairwise tests. In general, with the exception of Chinese vs. Danes (not significant) and the Chinese vs. Surui (significant) comparisons, Asian vs. Asian, Asian vs. New World and New World vs. New World comparisons were the only non-significant ones. These reduced 'geographical' differences in the distribution of maximum CTG/CAG repeat lengths reflect very well the known genetic similarities of populations (9–11). Rough estimates of the relative 'haploid' ligation-number frequencies in the populations analyzed were computed using the HAPLO program (12), assuming a linear dominance mode of inheritance (Table 1). Derived diversity (13) values were calculated as: 33% for Danes, 76% for Chinese, 55% for Japanese, 66% for Rondonian Surui, 71% for Maya and 54% for Mbuti/Biaka pygmies. These diversity values correspond to the probability of two gametes chosen at random having different maximum lengths as measured by number of $(CTG)_{17}$ oligos ligating. This is analogous to heterozygosity at an autosomal locus.

DISCUSSION

Our total genome scan for maximum CTG/CAG maximum repeat lengths demonstrates that RED with a $(CTG)_{17}$ is a reliable and highly reproducible method. While additional experiments would be required before any broad generalization can be strongly supported, we do find that even the relative intensities of ligation products within a sample show surprising reproducibility. Comparison of the results of three separate RED analyses done several months apart (Fig. 1) shows this reproducibility. We can only speculate, but intensity may reflect the number of copies per genome of sequences of the various longer lengths and the exact lengths of the segments. For example, a locus with a $(CTG)_{115}$ should more frequently (in a higher fraction of cycles) allow the final ligations of six $(CTG)_{17}$ than would a locus with a $(CTG)_{102}$.

Table 1. Estimated 'gametic' relative frequencies of maximum ligated (CTG)₁₇ for the populations analysed

	Danes (d = 33%)	Chinese (d = 76%)	Japanese (d = 55%)	R. Surui (d = 66%)	Maya (d = 71%)	Pygmies (d = 54%)
Maximum CTG ₁₇ size						
x = 2	0.81	0.32	0.62	0.35	0.32	0.61
x = 3	0.05	0.29	0	0.26	0.3	0.27
x = 4	0.07	0.15	0.25	0.38	0.29	0.1
x = 5	0.012	0.16	0.1	0	0.074	0
x = 6	0.012	0.075	0	0	0	0
x = 7	0.012	0	0	0	0	0
x = 8 or more	0.011	0	0.019	0	0	0

Based on a model of linear dominance these frequencies are estimations of the distributions of the maximum length in the gamete pools of the populations. Diversity values (d) for each population were calculated according to the formula:

$$d = 1 - \sum_{i=2}^n p(i)^2$$

where p(i) is the estimated gametic frequency of i ligated oligos and n is the maximum number seen in a population.

A striking result of this study is the large proportion of individuals with moderately long CTG/CAG repeats and the evidence that profound differences exist in the distributions of longest (CTG)_n repeat lengths in different human populations. The previous studies of Europeans had indicated that (CTG)_n greater than 84 repeats were rare (5–8) and even lengths between 51 and 84 repeats occurred in only a few individuals. RED results in the range of >67 repeats have often been referred to as 'expansions' (8,14). In contrast we find lengths up to 85–92 CTG/CAG repeats to be common in east Asian populations. While this size range corresponds to unstable disease-causing alleles at all 'CTG' or 'CAG' disease loci described so far (1–4), these lengths at the loci being detected in this study represent normal variation and would appear to be relatively stable. Thus, we cannot refer to these as 'expansions' with any sense of abnormality or dynamic change in length.

The most likely explanation for these observations is random genetic drift of alleles at (CTG)_n-containing loci. Additionally, we note that the mutation rate at such loci cannot be very great; a high mutation rate would most likely homogenize allele frequencies within a short evolutionary time, erasing any evidence of random genetic drift causing frequency differences among populations. In this light, gross instability events observed within populations must be considered a rather rare phenomenon, even in populations where repeat lengths of 78 CTG/CAG triplets or greater are frequent (Chinese, Japanese, Maya; Fig. 2). The (CTG)_n repeat sizes that we have observed within the Danish population correlate very well with the RED survey of (CAG)₁₇ maximum repeat lengths within northern Europeans reported by others (5–8). The very similar frequency pattern of maximum CTG/CAG repeat lengths observed by independent RED surveys in different northern European population samples minimizes the possibility that the CTG/CAG repeat lengths we detected may have arisen as new alleles from somatic mutations (as either contractions or expansions of CTG/CAG repeats) in the lymphoblastoid cell lines. While mutations at microsatellite loci in EBV transformed cell lines have been shown to occur, they are seen only after multiple passages in culture (all cell lines used in

our RED survey went through only a single passage and are overwhelmingly polyclonal) and, in the great majority of cases, involve only small length changes which would remain largely undetected by RED (15,16). This lack of fine tuning is due to an inherent methodological limit in the RED technique, where differences of maximum repeat lengths between samples can only be detected as multiples of the length of the oligonucleotide used in the reactions.

The differences among the non-human primates are not statistically significant, in large part because sample sizes are small, but some conclusions are possible. Both species of *Pan* are definitely polymorphic with some individuals having (CTG)_n sizes longer than seen in much larger samples of some human populations. Thus, we cannot say that repeats in humans are longer on average than in common and pygmy chimpanzees. This contrasts with the conclusion reached in a comparison of individual loci between humans and chimpanzees (17). However, because RED provides a genome-wide survey, the loci with long repeats in common and pygmy chimpanzees may be completely different from those in humans. That would reconcile the data sets because a significant advantage of our study, with respect to some previous studies on evolutionary dynamics of trinucleotide repeats in primates (e.g., ref. 17), is that the different levels of polymorphism and maximum lengths detected by RED at CTG/CAG trinucleotide repeat loci are not affected by ascertainment bias due to systematic selection for loci known *ab origine* to be polymorphic in humans. In contrast, both gorillas and orangutans appear to be less polymorphic than either *Pan* species (probability of no observed variation if the frequencies were as for *P. troglodytes* < 0.05) and certainly less polymorphic than most human populations. The most common repeat length in all non-human primates corresponds to ligation of three (CTG)₁₇ oligos reflecting genome lengths of (CTG)₅₁ to (CTG)₆₇. Humans differ by population: most individuals in east Asian and Amerindian populations have lengths greater than this while most individuals in European populations have smaller lengths. It seems likely that maximum lengths in the ancestral hominoid

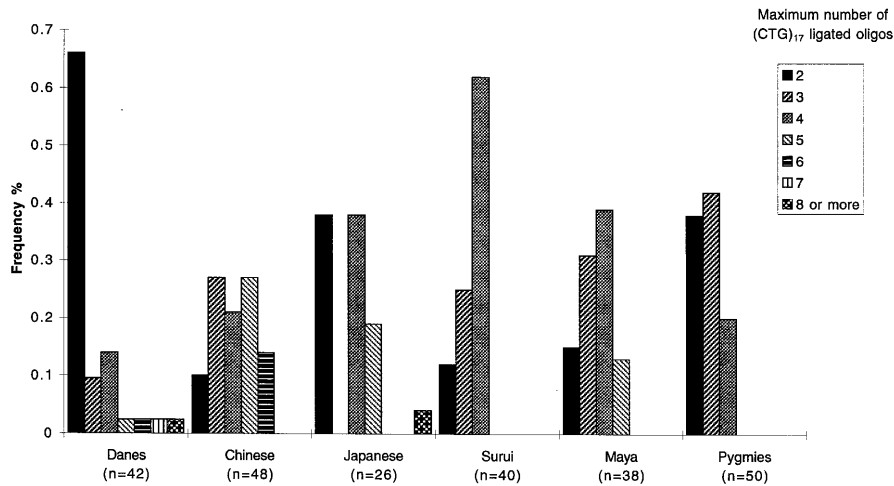


Figure 2. Results of our RED survey examining individuals from several different human populations. The distribution of the largest ligation product seen in individuals from different human populations is shown as the maximum number of (CTG)₁₇ oligos on the X axis, and the frequency of that size product among the individuals analyzed on the Y axis.

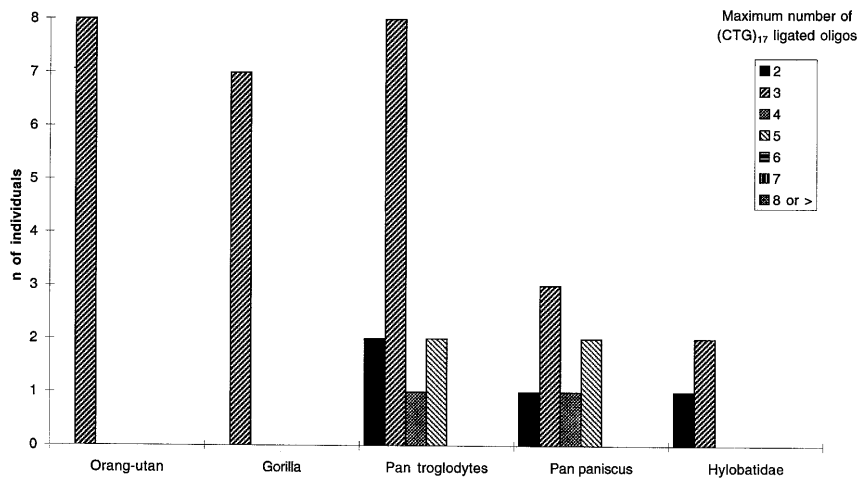


Figure 3. Distribution of ligated (CTG)₁₇ oligos in non-human primates. The number of (CTG)₁₇ ligation events is on the X axis, the absolute frequency of the individuals analyzed is on the Y axis (n = number of individuals analyzed).

were in the (CTG)₅₁₋₆₇ range with some gains and some losses as the modern species evolved.

In order to compute the gametic frequencies and the diversity values of the polymorphic CTG/CAG repeats detected by RED, we reasoned that the results obtained for a diploid individual would be identical to the results obtained for whichever of the two gametes that produced that individual had the largest CTG repeat in its haploid genome. Therefore, we can consider the phenotype a dominant phenotype related to an entire haploid genome. Assuming Hardy-Weinberg ratios, the frequency (f) of the phenotype with a longest ligation product of two (CTG)₁₇ repeats (r) would be equal to $f(2r) = p(2r)^2 + p(2r) \cdot p(1r) \cdot 2$, where p(x) is the frequency of gametes in which the longest stretch of (CTG)_n allows x (CTG)₁₇ oligos to ligate. This is identical to a linear dominance model where the HAPLO program can estimate the p(xr) values using the EM algorithm and the p(xr) values can then be used to estimate diversity. The most notable aspect of the data in Table 1 is the unusual distribution in the Danish sample resulting in a low diversity value. This approach is of course limited by the

very nature of the RED method, which does not differentiate between repeat containing loci, a fact which could obscure the correlation between ligation products and diversity values. However, while 'short' ligation ladders (up to five ligated fragments) most likely correspond to a great locus diversity, longer ligation ladders most probably correspond to fewer loci and may even represent a single locus at the longer fragments. Although calculated diversity estimates correspond only to a rough value of the actual diversity of CTG/CAG repeats in a given population, they nevertheless provide a useful measure from which to detect differences between well-defined populations.

These results also provide a degree of insight into the nature of (CTG)_n repeat loci and their stability over time. For example, a high frequency of individuals within any population (e.g. Chinese) displaying ligation products corresponding to multiple ligated (CTG)₁₇ suggests that a relatively high frequency of normal individuals homozygous for long CTG/CAG repeats may exist at a given locus and that no major selective role is acting against such long repeats in these populations. Consequently, we

can conclude that these long repeats cannot be a major cause of disease. Additionally, similarity of length distribution across the Americas, as shown by the same unimodal distribution observed in the two Amerindian populations (Maya and Surui; Fig. 2), suggests that loci with repeat lengths between (CTG)₅₁ and (CTG)₆₈ were present at a relatively high frequency in the original founding population of both Americas. Although these may be considered long repeat lengths relative to common lengths observed in Europeans and as such may be thought of as generated through pronounced instability at CTG/CAG-containing loci [involving gain or loss of multiple (CAG)₁₇], they appear to be quite stable in these populations. This interpretation however is plausible only under the assumption that the maximum repeat length of (CTG)₅₁ and (CTG)₆₈ detected by RED in independent Amerindian samples corresponds to the same locus or group of loci. The cloning and characterization of these CTG/CAG repeats will be necessary to confirm or refute such a hypothesis.

Some RED-based association studies of neuropsychiatric disorders in Europeans have recently suggested a possible pathogenic association of (CTG)_n repeats (mostly in the 180–255 bp range) with schizophrenia and bipolar affective disorder (6–8). In our RED survey CTG/CAG repeats of 180–210 bp belong to the groups with a maximum of three and four ligated (CTG)₁₇ oligos (Fig. 2). Individuals with these sizes were observed to be moderately to highly frequent in all populations analyzed, including the Danish population sample. Indeed, even larger sizes (five and six ligated oligos) are seen in one fifth to one third of the normal individuals from east Asia. Thus, interpreting RED results of (CTG)_{68–102} as pathogenic, even when found in a patient, is very problematic.

Given the complexity of the distribution of CTG/CAG maximum repeat lengths in populations, our survey findings indicate that in order to minimize population stratification artefacts, RED association studies with disease should always be performed by collecting accurate information on the normal variation of the triplet repeats of interest within the population studied. Our current survey serves to document great differences among populations but does not allow much basis for generalization to other populations. The only generalization that seems warranted is based on the consistent findings for samples of normal individuals from northern European populations (5–8): Europeans have a higher frequency of '2' and lower frequencies of '3' and '4' (CTG)₁₇ than populations in any other part of the world. In general the term 'expansion,' which implies instability and/or abnormality, does not seem to be appropriate for labeling the CTG/CAG containing loci detected by RED in our (or any) sample of unrelated normal individuals, at least in the range up to (CTG)₁₀₂, since these RED products could correspond to long but stable repeats, not prone to any dynamic change, unrelated to any expanding repeat per se, and not associated with any pathology. On the other hand, long (and rare) RED ligation products of (CTG)₁₁₉ and longer are much more likely to represent true dynamic mutations, and the term 'expansion' may be appropriate for them even in the absence of specific information on the locus corresponding to the repeat. Our findings suggest that, in general, lengths of trinucleotide repeats do not necessarily correlate with degree of instability, at least not in a simple linear way in the range up to (CTG)₆₈. The evidence from disease-causing loci does not appear to generalize to the majority of trinucleotide repeat loci elsewhere in the genome.

MATERIALS AND METHODS

Human population samples

All the human DNAs used in this study were purified from cultured cells using standard proteinase, phenol-chloroform extraction and alcohol precipitation (18). In total 244 individuals were tested. The Japanese (N = 26) and Pygmy (N = 50) samples were collected and described by Cavalli-Sforza *et al.* (1986) (19). The Japanese sampled are from the San Francisco area or students/postdocs from Japan and the Pygmies are from the Central African Republic (Biaka) and Zaire (Mbuti). The Chinese sample (N = 48) was collected in Taiwan by R.B. Lu (20). The Danish sample (N = 42) was collected by J. Parnas and consists of unrelated individuals from the Copenhagen area. Rondonian Surui (N = 40) were collected by F.L. Black in the Rondonia province, western Amazon basin. The Maya (N = 38) were collected in the Yucatan peninsula by K. Weiss. Both the R. Surui and Maya were described by Kidd *et al.* (1991) (21).

Non human primate samples

All DNA samples of non-human primates were extracted from lymphoblastoid cell lines established in our laboratory unless otherwise noted. The same protocol for DNA extraction was used as for the human DNA samples (18). The provenances of non-human primates sampled for this project were as follows:

(i) Orang-utans (*Pongo pygmaeus*): Sibü and Tupa were obtained from the Yerkes Regional Primate Center, Atlanta, GA; Ben was from the Henry Doorly Zoo, Omaha, NE; cell lines from Jari, Puti and CP81 were obtained from David Lawlor of Stanford University; Sunda was from the Sacramento Zoo, Sacramento, CA; (ii) Gorillas (*Gorilla gorilla*): Oko, Cal, and Ozoum were from the Yerkes Regional Primate Center; Abe and Murphy were from the Henry Doorly Zoo; cell lines from Rok and Machi were obtained from David Lawlor; (iii) Common chimpanzees (*Pan troglodytes*): Harriet was from the Arizona Primate Foundation, Tempe, AZ; Herman was from the Lowery Zoo, Tampa, FL; Hannibal, Juno, Bullet, and Lottie were from Alfred Prince, New York Blood Center, New York, NY; Bakoumba, Cheetah, Mabolite, and Julie were from the Centre International de Recherches Medicales de Franceville; A333, A336, and A208 were from the New Iberia Research Center, New Iberia, LA; (iv) Pygmy chimpanzees (*Pan paniscus*): Kidongo, Matata, Linda and Bosondjo were from the Yerkes Regional Primate Center; Lady and Maringa were from the Milwaukee County Zoo, Milwaukee, WI; (v) *Hylobatidae* (*Hylobates lar*, *Hylobates klossii* and *Hylobates syndactylus*): lymphoblastoid cell lines from a single *Hylobates lar*, a single *Hylobates klossii* and a single *Hylobates syndactylus* were obtained from David Ward of Yale University.

RED

In the RED method Ampligase, thermostable DNA ligase (Epicentre Technologies, Madison, WI) is used in a cycling procedure that generates multimers of the oligonucleotide(s) utilized. Following cycling, the reaction products are electrophoresed, blotted and detected by hybridization with a radiolabeled oligonucleotide probe complementary to the oligonucleotide used in the ligation reactions. We have modified the original RED protocol described by Schalling *et al.* (5,22). We used ~5 µg of genomic DNA, a (CTG)₁₇ oligo and 5 U Ampligase in all

reactions (20 λ). We observed a linear increase of the signal strength when five μg of genomic DNA (2.5 attomoles/ λ) were used, instead of 1 μg (0.5 attomoles/ λ) as originally described (5). By using a (CAG)₁₄₁ cloned in Bluescript as template, Schalling *et al.* (5) have previously shown that increasing amounts of template DNA, ranging from 0.5 to 5 attomoles of repeat / λ (1 to 10 pg of plasmid DNA) in the ligation reaction do not have the effect of producing artefactual multimers but improve the strength of the hybridization signal. Faint artefactual bands were only observed when the plasmid DNA template was increased to 50 attomoles/ λ per reaction (100 pg) (5). The size of the largest ligation product is highly repeatable (Fig. 1A and B) and varies among individuals. After 400 cycles (95°C, 10 s; 70°C 30 s), the reaction products were fractionated by electrophoresis on a 6% acrylamide denaturing gel and transferred onto a nylon membrane. Multimers were then detected by hybridization with a radio-labeled oligonucleotide (CAG)₁₀ probe.

HAPLO program

Gametic frequency estimates were made using the HAPLO program (12). Program and documentation are available via anonymous FTP from paella.med.yale.edu, in the directory pub/haplo.

ACKNOWLEDGMENTS

G.S. dedicates this paper to the memory of Elisabetta Sirugo. We thank Drs Francesc Calafell and Wayne Fenton and Matthew Hawley for helpful discussions and advice. This work was supported by research grants MH-30929, MH-39239, MH-50390 (to K.K.K.), NSF Grant SBR 9408934 (to J.R.K.) and by grants of the Leakey Foundation and Wenner-Gren Foundation for Anthropological Research (to A.S.D.).

REFERENCES

1. Richards R.I. and Sutherland G.R. (1992) *Nature Genet.*, **1**, 7–9.
2. Warren S.T. (1996) *Science*, **271**, 1374–1375.
3. Zoghbi H.Y. (1996) *Nature Genet.*, **14**, 237–238.
4. Harris S., Moncrieff C. and Johnson K. (1996) *Hum. Mol. Genet.*, **5**, 1417–1423.
5. Schalling M., Hudson T.J., Buetow K.H. and Housman D.E. (1993) *Nature Genet.*, **4**, 135–139.
6. O'Donovan M.C., Guy C., Craddock N., Murphy K.C., Cardno A.G., Jones L.A., Owen M.J. and McGuffin P. (1995) *Nature Genet.*, **10**, 380–381.
7. Lindblad K., Nylander P.O., De Bruyn A., Sourey D., Zander C., Engstrom C., Holmgren G., Hudson T., Chotui J., Mendlewicz J., Broeckhoven C. and Schalling M. (1995) *Neurobiol. Disease*, **2**, 55–62.
8. Morris A.G., Gaitonde E, Mc Kenna P.J., Mollon J.D. and Hunt D.M. (1995) *Hum. Mol. Genet.*, **4**, 1957–1961.
9. Cavalli-Sforza L.L., Menozzi P., Piazza A. (1995) *The history and geography of human genes*. Princeton University press.
10. Kidd K.K. and Kidd J.R. (1996) In Boyce, A.J. and Mascie-Taylor C.G.N. (Eds), *Molecular biology and human diversity*, Cambridge University Press.
11. Tishkoff S.A., Dietzsch E., Speed W.C., Pakstis A.J., Cheung K., Kidd J.R., Bonne-Tamir B., Santachiara-Benerecetti A.S., Moral P., Watson E., Krings M., Paabo S., Risch N., Jenkins T. and Kidd K.K. (1996) *Science*, **271**, 1380–1387.
12. Hawley M.E. and Kidd K.K. (1995) *J. Hered.*, **86**, 409–411
13. Nei M. (1977) *Ann. Hum. Genet.*, **41**, 225–239.
14. Bleyl S., Nelson L., Odelberg S. J., Rutenberg H.D., Otterud B., Leppert M. and Ward K. (1995) *Am. J. Hum. Genet.*, **56**, 408–415.
15. Banchs I., Bosch A., Guimera J., Lazaro C., Puig A. and Estivill X. (1994) *Hum. Mutat.*, **3**, 365–372.
16. Ashizawa T., Monckton D.G., Vaishnav A., Patel B.J., Voskova A. and Caskey C.T. (1996) *Genomics*, **36**, 47–53.
17. Rubinsztein D.C., Amos W., Leggo J., Goodburn S., Jain S., Li S-H., Margolis R.L., Ross C.A. and Ferguson-Smith M.A. (1995) *Nature Genet.*, **10**, 337–343.
18. Sambrook J., Fritsch E.F., Maniatis T. (1989) *Molecular Cloning: a laboratory manual, 2nd edn.*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
19. Cavalli-Sforza L.L., Kidd J.R., Kidd K.K., Bucci C., Bowcock A.M., Hewlett B.S. and Freidlaender J.S. (1986) *Cold Spring Harbor Symposia on Quantitative Biology*, **51**, 411–417.
20. Lu R-B., Ko H-C., Chang F-M., Castiglione C.M., Schoolfield G., Pakstis A.J., Kidd J.R. and Kidd K.K. (1996) *Biol. Psychiatry*, **39**, 419–429.
21. Kidd J.R., Black F.L., Weiss K.M., Balazs I. and Kidd K.K. (1991) *Human Biology*, **63**, 775–794.
22. Sirugo G. and Kidd K.K. (1995) *Epicentre Forum*, **2**, 1–3.