Patterns of Instability of Expanded CAG Repeats at the *ERDA1* Locus in General Populations

Ranjan Deka,[1] Sun Guangyun,[1,3] Jonathan Wiest,[1] Diane Smelser,[1] Su Chunhua,[1] Yixi Zhong,[2] and Ranajit Chakraborty[2]

[1]Department of Environmental Health, University of Cincinnati, Cincinnati; [2]Human Genetics Center, School of Public Health, University of Texas Houston Health Science Center, Houston; and [3]Department of Forensic Biology, West China University of Medical Sciences, Chengdu, People's Republic of China

Summary

A highly polymorphic CAG repeat locus, *ERDA1*, was recently described on human chromosome 17q21.3, with alleles as large as 50–90 repeats and without any disease association in the general population. We have studied allelic distribution at this locus in five human populations and have characterized the mutational patterns by direct observation of 731 meioses. The data show that large alleles (⩾40 CAG repeats) are generally most common in Asian populations, less common in populations of European ancestry, and least common among Africans. We have observed a high intergenerational instability (46.3% ± 5.1%) of the large alleles. Although the mutation rate is not dependent on parental sex, paternal transmissions have predominantly resulted in contractions, whereas maternal transmissions have yielded expansions. Within this class of large alleles, the mutation rate increases concomitantly with increasing allele size, but the magnitude of repeat size change does not depend on the size of the progenitor allele. Sequencing of specific alleles reveals that the intermediate-sized alleles (30–40 repeats) have CAT/CAC interruptions within the CAG-repeat array. These results indicate that expansion and instability of trinucleotide repeats are not exclusively disease-associated phenomena. The implications of the existence of massively expanded alleles in the general populations are not yet understood.

Received January 8, 1999; accepted for publication April 27, 1999; electronically published May 25, 1999.

Address for correspondence and reprints: Dr. Ranjan Deka, Department of Environmental Health, University of Cincinnati, 3223 Eden Avenue, Cincinnati, OH 45267-0056. E-mail: ranjan.deka@uc.edu

0002-9297/99/6501-0025$02.00

Introduction

Trinucleotide repeats have been implicated in the genesis of more than a dozen human diseases. Of these, the CAG repeats alone are shown to be the cause of at least eight neurodegenerative disorders, including Kennedy disease (MIM 313200; LaSpada et al. 1991), Huntington disease (MIM 143100; The Huntington's Disease Collaborative Group 1993), dentatorubral pallidoluysian atrophy (MIM 125370; Koide et al. 1994), and several types of spinocerebellar ataxias: SCA1 (MIM 164400; Orr et al. 1993), SCA2 (MIM 183090; Pulst et al. 1996), SCA3 (MIM 109150; Kawaguchi et al. 1994), SCA6 (MIM 183086; Zhuchenko et al. 1997), and SCA 7 (MIM 164500; David et al. 1997). All these repeat loci are, however, moderately to highly polymorphic in normal individuals. Disease results when the repeat arrays become longer than a certain threshold. In the case of CAG-repeat–associated diseases, the longest normal allele does not exceed 40 repeats. Recently, two polymorphic CAG/CTG repeat loci, *ERDA1* (Nakamoto et al. 1997) and *SEF2-1* (Breschel et al. 1997), have been reported with massively expanded alleles in general populations. Neither of these two loci is shown to be associated with any known disease. The expanded alleles at these two loci appear to confound the repeat expansion–detection (RED) method used for searching disease-causing loci with trinucleotide-repeat expansion (Benson et al. 1998). The *ERDA1* (expanded repeat domain CAG/CTG 1) locus, with alleles as large as 50–90 repeats, is located on human chromosome 17q21.3. Ikeuchi et al. (1998) describe apparently the same locus with the designation *Dir1*, demonstrating intergenerational instability of the large alleles (>50 CAG repeats). We have studied allelic distribution at the *ERDA1* locus in five human populations and have characterized the mutational patterns by direct observation of 731 meiosis events. Our data show that large CAG-repeat alleles (⩾40 repeats) are generally most common in Asian populations, less common in populations of European ancestry, and least common in African populations. We have observed a high intergenerational instability

(46.3% ± 5.1%) of the large alleles, with a contraction bias in male and an expansion bias in female meioses. Within this class of large alleles, mutation rate increases concomitantly with increasing allele size. However, the magnitude of the expansion or contraction of the mutated allele does not appear to depend on the size of the progenitor allele.

Material and Methods

DNA Samples, PCR Analysis, and Sequencing

The population samples (German, Nigerian, New Guinea Highlander, and Chinese) have been described elsewhere (Deka et al. 1995). The southwestern Chinese families were drawn from the Sichuan province of the People's Republic of China. These samples were typed at 14 highly polymorphic loci for paternity exclusion. Amplification of the CAG repeats was performed according to the protocol of Nakamoto et al. (1997). For DNA sequencing, PCR products were gel purified by means of the Qiagen gel extraction kit QIAEXII and were sequenced with the BigDye terminator cycle sequencing ready reaction kit from PE Applied Biosystems. Both strands of DNA were sequenced on an ABI 377 sequencer with the forward and reverse primers as described by Nakamoto et al. (1997).

Data Analysis

Allele frequencies were estimated by the gene count method (Li 1976). Allele size variance (henceforth referred to as "variance") and heterozygosity were computed with the estimating equations given in Kimmel et al. (1996) and Nei (1978), respectively. Agreement with Hardy-Weinberg equilibrium (HWE) of genotype frequencies was tested by following the procedure described in Guo and Thompson (1992), with 10,000 replications of permutation of alleles within each sample. Mutational count data were treated as binomial proportions, so that standard errors (SEs) of all estimates were computed by the standard theory of proportions. Allele size dependency of mutation rates (see fig. 3) was evaluated by the test for a linear trend of proportions (Snedecor and Cochran 1976). Homogeneity of allele frequencies across samples was tested through the r × c contingency table analysis, in which the levels of significance were determined by 10,000 permutations of alleles. The maximum-likelihood method (Chakraborty et al. 1996) was used for testing segregation distortion of transmission of short alleles from the family data.

Results

Allelic Distributions in Human Populations

Using the PCR protocol described by Nakamoto et al. (1997), we analyzed DNA samples from unrelated individuals (42 Han Chinese, 49 Germans, 51 Nigerians, and 49 Papua New Guinea Highlanders) with no known diseases. In addition, we studied 38 CEPH panel pedigrees and 43 Chinese nuclear families (from southwestern China), which provided information on a total of 731 meioses and genotypic data on 113 unrelated individuals from the CEPH cohort and on an additional 85 unrelated Chinese individuals. Allele frequency estimates from the two Chinese population samples were not significantly different ($P = .38$); therefore, for subsequent analyses, these two samples were combined. Figure 1 shows the frequency distributions of alleles (designated by CAG-repeat sizes) in the five populations (Chinese, New Guinean, German, unrelated CEPH individuals, and Nigerian). In general, although allele size

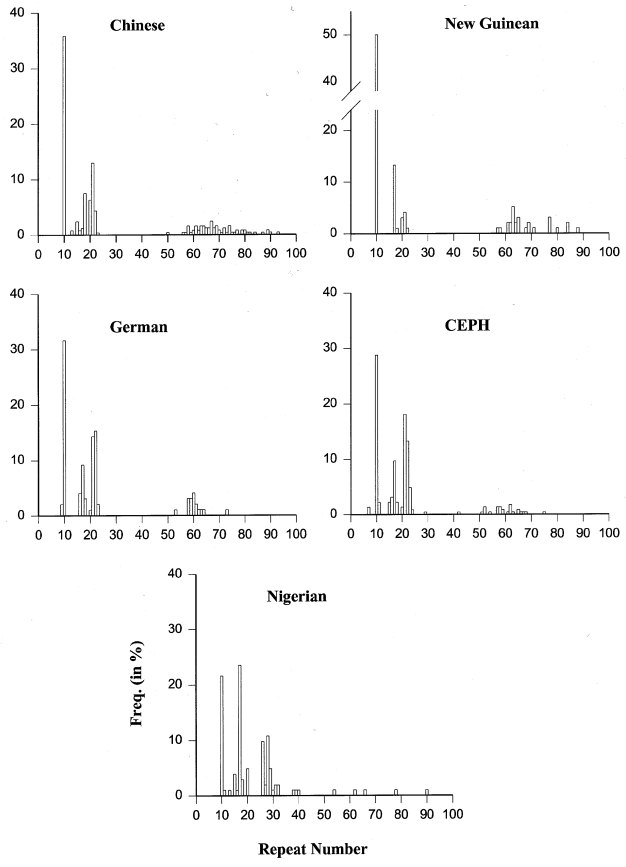

Figure 1 Distribution of the CAG/CTG–repeat alleles at the *ERDA1* locus in five human populations. The number of chromosomes sampled from each population from which the gene-count estimates of allele frequencies were computed is shown in table 1.

distributions are bimodal in all populations, as reported elsewhere (Nakamoto et al. 1997; Ikeuchi et al. 1998), our data show that the large alleles (≥40 repeats) are relatively less common (5.9%) in the Nigerian population and that the intermediate-sized alleles (27–40 repeats), which are almost absent elsewhere, are quite common (25.5%) among Nigerians. Furthermore, although repeat 10 is the most common allele in all other populations, among the Nigerians the 17-repeat allele is at least as common (23.5%) as the 10-repeat allele (21.6%). We also observed a few alleles containing <10 repeats in both populations of European ancestry (German and CEPH), which had not been reported in the above-mentioned two studies (Nakamoto et al. 1997; Ikeuchi et al. 1998).

A summary of the population genetic characteristics of the locus is presented in table 1. These data indicate that *ERDA1* is a highly polymorphic locus with a heterozygosity of 73%–88%. Relatively lower heterozygosity (73%) in New Guineans, as observed at other microsatellites (Deka et al. 1995), is a reflection of their evolutionary history and relative isolation. Interestingly, however, they have the highest frequency of the large alleles (27.6%), some of which are quite large (88 repeats). Consequently, we observe the largest variance in this population. In fact, unlike at other microsatellite loci (Chakraborty et al. 1997), variance and average heterozygosity, at this locus, are not positively correlated, although both are measures of within-population genetic variability. For example, variance is significantly lower in German, CEPH, and Nigerian, in spite of their higher heterozygosity. This imbalance between variance and heterozygosity is apparently due to the atypical allele size distribution at this locus, where two classes of alleles are cosegregating, with perhaps different dynamics of mutational patterns. With respect to concordance of genotype frequencies with Hardy-Weinberg expectations, the German and Nigerian samples show significant (at 5% level) deviation, with the CEPH sample being in marginal agreement ($P = .052$). It should be noted that the large alleles are relatively infrequent in these three populations (see table 1).

Mutation at the ERDA1 Locus

Since expanded alleles at disease-causing trinucleotide repeats are meiotically unstable, we examined whether the large alleles at *ERDA1* also show intergenerational instability. Ikeuchi et al. (1998) had reported instability of large alleles at the *Dir1* locus in a smaller number ($n = 69$) of meioses. From 38 multigenerational CEPH panel pedigrees, we obtained data on parent-offspring transmission of 315 paternal and 325 maternal alleles, of which 43 paternal and 18 maternal transfers involved alleles of ≥40 repeats. Similarly, in the sample of 43 southwestern Chinese nuclear families, we observed intergenerational transmission of alleles in 45 paternal and 46 maternal meioses, in which 18 paternal and 16 maternal alleles were ≥40 repeats. In total, we observed 45 mutations in 731 meioses (a mutation rate of 6.2% ± 0.9% per generation). Only one of these mutations involved an allele of <40 repeats (a maternal 17-repeat allele had a single-repeat contraction in the CEPH panel). In the total data, although paternal mutations (28 of 360 meioses) appear to be more common (7.8% ± 1.4%) than maternal mutations (17 of 371 meioses; 4.6% ± 1.1%), this difference is not statistically significant ($P = .093$). Furthermore, when the transmissions of only the large-sized alleles are considered, mutation rates (45.9% ± 6.4% paternal and 47.1% ± 8.6% maternal) do not show any bias with respect to parental origins.

Although not shown in detail, a higher mutation rate was observed in the Chinese families (22% ± 4.3%) than in the CEPH families (3.9% ± .8%). This is, however, an artifact caused by allele frequency differences in these two populations. Expanded alleles (≥40 repeats) are more frequent in the Chinese population (27.5%) than in the CEPH samples (11.5%). Furthermore, in the Chinese population, the alleles of ≥61 repeats account for most of the expanded alleles (86%), whereas in the CEPH panel sample only 46% of the expanded alleles are >60 repeats.

Mutation patterns were, however, different, depending on the sex of parents. Paternal transmissions pre-

Table 1

Population Genetic Characteristics of CAG/CTG-Repeat Polymorphism at the *ERDA1* Locus in Five Human Populations

Population	Sample Size[a]	Repeat Size	Frequency ± SE of Large Alleles	Variance	Expected Heterozygosity	Level of Significance of HWE Test
Chinese	254	10–93	27.5 ± 2.8	631.4	.843	.678
New Guinean	98	10–88	27.6 ± 4.5	655.9	.730	.689
German	98	9–73	17.4 ± 3.8	313.4	.848	.031
CEPH	226	7–75	11.5 ± 2.1	218.9	.855	.052
Nigerian	102	10–90	5.9 ± 2.3	185.4	.876	.020

[a] Sample sizes are the number of chromosomes studied from unrelated individuals in each population.

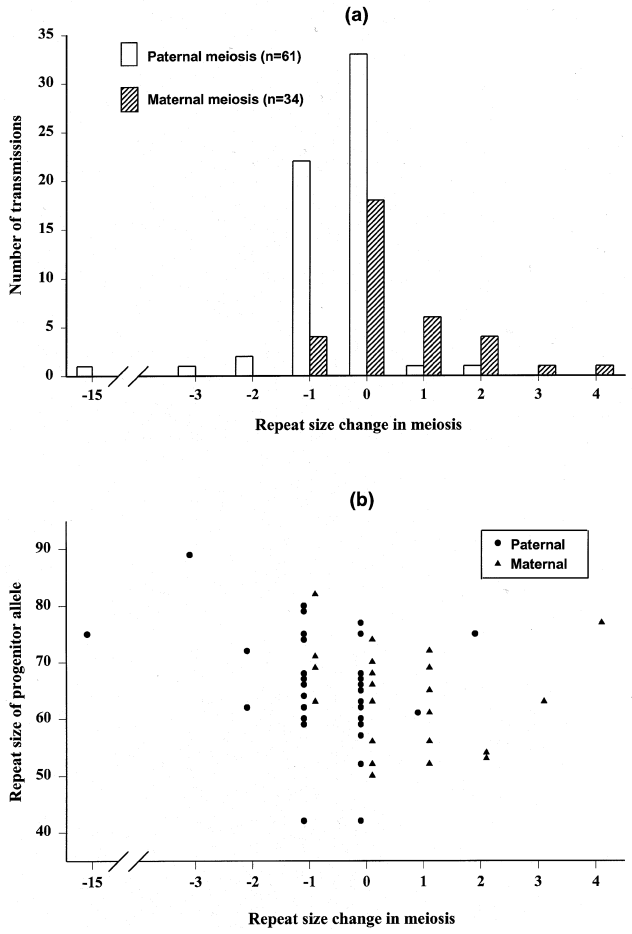

Figure 2 *a,* Distribution of intergenerational changes of ⩾40 CAG/CTG repeats at the *ERDA1* locus during paternal and maternal meioses. *b,* Distribution of progenitor alleles (vertical axis) plotted against the size change produced by these alleles during transmission (horizontal axis). Meioses showing no change in length are represented as zero change in repeat sizes. Note: the number of data points in *b* is smaller than the total number of meioses ($n = 91$) because some progenitor alleles produced multiple identical-length changes.

dominantly resulted in contraction of allele sizes (with a mean reduction of 0.67 ± 0.26 repeats in 61 transmissions), whereas most maternal transmissions resulted in expansions (with a mean increase of 0.50 ± 0.19 repeats in 34 transmissions, excluding the single mutation that involved a 17-repeat allele). The observed distributions of allele size changes are shown in figure 2*a*. As seen from this figure, a single large (15-repeat) contraction mutation was observed in a CEPH male individual involving a progenitor allele of 75 repeats. Of the 28 paternal mutations, 26 (93%) resulted in contractions, whereas, of the 16 maternal mutations (⩾40 repeats), 12 (75%) resulted in expansions. Qualitatively, these results are consistent with the findings of Ikeuchi et al. (1998), with the exception that, in our larger sample (95 transmissions of large alleles, as opposed to 26 studied by those authors), the rate of instability of large alleles was not found to be dependent on parental origin (28/61 in paternal as opposed to 16/34 in maternal transmissions).

The larger number of meioses allowed us to further examine whether, within the large-sized alleles, the magnitude of contraction and expansion of the mutated alleles is dependent on the size of the progenitor allele. In figure 2*b,* we have plotted the size of the progenitor alleles (paternal and maternal) against the size change in the mutated allele during meiosis. Progenitor allele sizes, which did not mutate during transmission, are also shown in the figure (repeat-size change "zero" in the horizontal axis). We found no correlation between length change of the mutated alleles and the length of the progenitor alleles in either sex. However, progenitor alleles showing no change in size during transmission are, as a group, somewhat smaller in size than the progenitors that have mutated. Paternal alleles ($n = 33$) without mutation have a mean size of 62.2 ± 1.2 repeats versus a mean of 67 ± 1.7 for alleles with mutation ($n = 28$). Similarly, mean repeat sizes of maternal alleles that have not mutated ($n = 18$) and that have mutated ($n = 16$) are 59.3 ± 1.9 and 63.3 ± 2.4, respectively. These differences are, nonetheless, not statistically significant in either sex (paternal $P = .18$; maternal $P = .08$).

Our data, however, show that mutation rate within the large alleles varies with the size of the progenitor allele. This finding is shown in figure 3, in which we grouped the parental alleles into three size ranges: <40 repeats, 40–60 repeats, and ⩾61 repeats. In maternal as well as paternal transmissions (and, consequently, in the

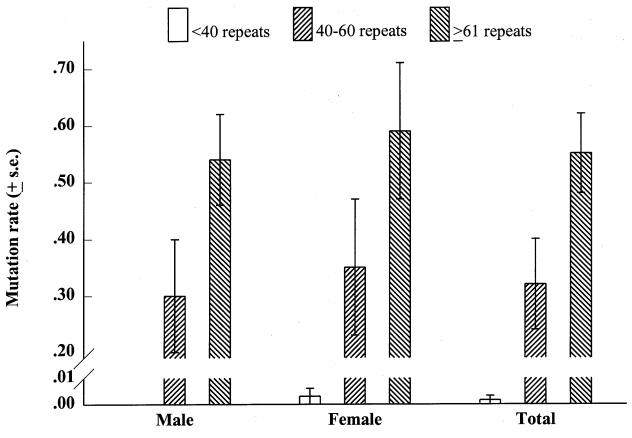

Figure 3 Allele size–specific (⩽40 repeats, 41–60 repeats, and ⩾61 repeats) mutation rates (with 1 SE error bar) at the *ERDA1* locus. Estimates were obtained separately for paternal and maternal transmissions as well as for the total sample.

total data), there is a significant positive trend of increase of mutation rates with increasing sizes of progenitor alleles ($P < .0005$ for all three tests).

Interruption within the CAG-Repeat Array

As mentioned earlier, the intermediate-sized CAG-repeat alleles are almost exclusively present in Nigerians, whereas the large alleles are uncommon (<6%). It is then of interest to determine whether any sequence heterogeneity exists in alleles of <40 repeats, particularly in the intermediate-sized alleles. We sequenced 18 small (10–26 repeats) alleles (4 from chimpanzees, 6 from the CEPH panel, 2 from Germans, 4 from Chinese, and 2 from Nigerians) and 5 intermediate-sized alleles (from the Nigerians). The sequences observed in these 23 alleles are shown in table 2. The flanking sequences of all the sequenced alleles are in conformity with the ones published by Nakamoto et al. (1997) and Ikeuchi et al. (1998). The small alleles, irrespective of their population origin, consist of uninterrupted CAG repeats. In contrast, the intermediate-sized alleles, from the Nigerians, have CAT and CAC interruptions within the CAG-repeat array. Of the five intermediate-sized alleles, three are of sizes 30, 31, and 32 repeats, with a $(CAG)_9(CAT)(CAG)_n$ motif. The interruption pattern is more complex for the other two intermediate-sized alleles, of 38 and 39 repeats, which have a $(CAG)_{13}(CAT)(CAG)(CAC)(CAG)_n$ motif.

Discussion

The above observations outline the implications of CAG/CTG repeat polymorphisms on the stability of the genome, even when they are not necessarily associated with clinical phenotypes. In general, alleles of <40 repeats are relatively stable. Our finding of one mutation in 636 meioses (i.e., a mutation rate of .0016 per generation) involving small alleles is consistent with the rate of mutations at other microsatellite loci (Weber and Wong 1993). In contrast, expanded alleles (⩾40 repeats) are primarily unstable. Within this class of large alleles, we also observed allele size dependency of mutation rates, with mutation rates reaching almost 55% for alleles of ⩾61 repeats. These observations mimic the phenomena of anticipation and dynamic mutation, which are features of disease-associated trinucleotide-repeat loci. Further, in the context of sex-of-origin effect on mutations, although mutation rate is apparently not dependent on parental sex, paternal mutations have predominantly resulted in contractions, whereas maternal mutations have yielded expansions. This relationship between contraction bias, on the one hand, and expansion bias, on the other, is perhaps an evolutionary balancing mechanism restricting the alleles from expanding indefinitely. Alternatively, segregation distortion could play a role in maintaining the allele sizes within the observed ranges. We investigated whether the high mutation rate of the expanded alleles is being countered by preferential transmission of small-sized alleles when one or both parents are heterozygous, with one allele being large and the other small. In our sample, there are 58 informative families with at least one heterozygous (with one short and one expanded allele) parent, which provided allele transmission data on a total of 154 children. Data were analyzed by a method described elsewhere (Chakraborty et al. 1996); however, we found no evidence of significant segregation distortion (estimates of the probability of transmission of short alleles were $.53 \pm .05$ of paternal alleles and $.44 \pm .07$ of maternal alleles, $P > .05$ for both when tested for the absence of segregation distortion). Thus, the evolutionary balance—that is, maintenance of allele sizes within the observed

Table 2

Sequence Motifs of Selected ERDA1 Alleles

Sample ID	Population Origin	Repeat Size	Sequence
Chimp1[a]	Chimpanzee	12	$G(A)_5G(CAG)_{12}(A)_5GAC$
Chimp2[a]	Chimpanzee	26	$G(A)_5G(CAG)_{26}(A)_5GAC$
CN 2[a], CN 4[a]	Chinese	10	$G(A)_5G(CAG)_{10}(A)_5GAC$
CEPH 133402[a]	CEPH	10	$G(A)_5G(CAG)_{10}(A)_5GAC$
CEPH 135003[a]	CEPH	17	$G(A)_5G(CAG)_{17}(A)_5GAC$
CEPH 6602[a]	CEPH	21	$G(A)_5G(CAG)_{21}(A)_5GAC$
GR 48[a]	German	22	$G(A)_5G(CAG)_{22}(A)_5GAC$
BE 10	Nigerian	15	$G(A)_5G(CAG)_{15}(A)_5GAC$
BE 14	Nigerian	18	$G(A)_5G(CAG)_{18}(A)_5GAC$
BE 68	Nigerian	30	$G(A)_5G(CAG)_9CAT(CAG)_{20}(A)_5GAC$
BE 14	Nigerian	31	$G(A)_5G(CAG)_9CAT(CAG)_{21}(A)_5GAC$
BE 15	Nigerian	32	$G(A)_5G(CAG)_9CAT(CAG)_{22}(A)_5GAC$
BE 27	Nigerian	38	$G(A)_5G(CAG)_{13}CAT(CAG)CAC(CAG)_{22}(A)_5GAC$
BE 10	Nigerian	39	$G(A)_5G(CAG)_{13}CAT(CAG)CAC(CAG)_{23}(A)_5GAC$

[a] Homozygous samples, accounting for sequences of two alleles each.

ranges—at the *ERDA1* locus appears to be dictated primarily by the sex-of-origin effect on mutational pattern.

Our observation of CAT/CAC interruption in the intermediate-sized alleles in the Nigerians (table 2) may relate to the interpopulation differences in frequencies of alleles of the different size classes and the evolution of the large-sized alleles. Recall that intermediate-sized alleles are almost exclusively present in Nigerians and are almost absent in the other four populations. Our data on mutations suggest that the small alleles are inherently stable. Interruptions in the intermediate-sized alleles probably preserve their stability as well. Sequence interruptions are known to regulate repeat stability. At least in two of the trinucleotide-repeat disorders—SCA1 and fragile X syndrome (MIM 309550)—sequence interruptions within the repeat array have been shown to confer stability to the normal-sized alleles (Chung et al. 1993; Eichler et al. 1994). Absence of interruption in small as well as large alleles (our data in table 2 and those of Nakamoto et al. 1997 and Ikeuchi et al. 1998) suggests that the intermediate-sized alleles in the Nigerians are not the source of the large unstable alleles. Nakamoto et al. (1997) found a single nucleotide polymorphism immediately flanking the CAG repeats at the 3′ end, with an "A" in short alleles and a "C" in long alleles in both Japanese and white individuals. Presence of the A residue in the polymorphic site in the intermediate-sized alleles in our Nigerian samples also suggests that these alleles are probably not the progenitors of the large alleles in the non-African populations. Whether the A→C substitution alone is responsible for the instability of the large alleles cannot be determined from the present data. However, flanking sequence polymorphism has been shown to regulate minisatellite mutation rates (Monckton et al. 1994). Thus, it may not be surprising if such mechanisms were also to regulate mutational processes in microsatellite loci.

We have also determined the CAG-repeat length in a sample of 10 chimpanzees. With a total of 10 segregating alleles, the chimpanzees are highly heterozygosity polymorphic, with the unbiased estimate of heterozygosity being 98%. However, all the alleles belong to the smaller class of repeats. As shown in table 2, the sequences of two alleles (sizes 12 and 26 repeats) have an uninterrupted CAG repeat followed by an A residue in the 3′ end, identical to those of the stable human alleles. This favors the smaller repeats being the ancestral state in humans.

Using northern blot and reverse transcription–PCR analysis, Ikeuchi et al. (1998) demonstrated that the *ERDA1/Dir1* repeat is transcribed in human brain and testis. However, it is not known whether the CAG-repeat region is located within an open reading frame coding for a polyglutamine tract. Nor has it been shown to be associated with any disease phenotype. It is also possible that the *ERDA1* repeat is present in the untranslated region of a gene and acts in the expression of a phenotype similar to the CTG repeats in both myotonic dystrophy (MIM 160900; Mahadevan et al. 1992) and the most recently reported spinocerebellar ataxia type, SCA8 (MIM 603680; Koob et al. 1999), and the GAA repeat in Friedreich ataxia (MIM 229300; Campuzano et al. 1996). However, no CAG repeat in an untranslated region has thus far been shown to have such a mechanism. The *SEF2-1* repeat, mentioned earlier, is also located within an intron of the *SEF2-1* gene (Breschel et al. 1997). The observation of expanded alleles with intergenerational instability at these transcribed repeats raises the possibility of the existence of many more trinucleotide-repeat loci, within and in close proximity to genes, with the characteristics of expansion and instability and without association with a disease phenotype. This would imply that the process of instability of repeat alleles is not exclusively a disease-associated phenomenon. Alternatively, these repeats are associated with unknown disease phenotypes, the implication of which has yet to be understood.

Acknowledgments

We thank Drs. C. H. Bunker, R. E. Ferrell, G. Flatz, and P. E. Smouse for providing various population DNA samples. This research is supported by U.S. Public Health Service Research grants GM 45861 and GM 41399.

Electronic-Database Information

Accession numbers and URL for data in this article are as follows:

Online Mendelian Inheritance in Man (OMIM): http://www.ncbi.nlm.nih.gov/Omim (for Kennedy disease [MIM 313200], Huntington disease [MIM 143100], dentatorubral pallidoluysian atrophy [MIM 125370], SCA1 [MIM 164400], SCA2 [MIM 183090], SCA3 [MIM 109150], SCA6 [MIM 183086], SCA7 [MIM 164500], SCA8 [MIM 603680], fragile X syndrome [MIM 309550], myotonic dystrophy [MIM 160900], and Friedreich ataxia [MIM 229300])

References

Benson KF, Horwitz M, Wolff J, Friend K, Thompson E, White S, Richards RI, et al (1998) CAG repeat expansion in autosomal dominant familial spastic paraparesis: novel expansion in a subset of patients. Hum Mol Genet 7:1779–1786

Breschel TS, McInnis MG, Margolis RL, Sirugo G, Corneliussen B, Simpson SG, McMahon FJ, et al (1997) A novel, heritable, expanding CTG repeat in an intron of the SEF2–1 gene on chromosome 18q21.1. Hum Mol Genet 11:1855–1863

Campuzano V, Montermini L, Moltò MD, Pianese L, Cossée M, Cavalcanti F, Monros E, et al (1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science 271:1423–1427

Chakraborty R, Kimmel M, Stivers DN, Davison LJ, Deka R (1997) Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. Proc Natl Acad Sci USA 94:1041–1046

Chakraborty R, Stivers DN, Deka R, Yu LM, Shriver MD, Ferrell RE (1996) Segregation distortion of the CTG repeats at the myotonic dystrophy locus. Am J Hum Genet 59: 109–118

Chung M-Y, Ranum LPW, Duvick LA, Servadio A, Zoghvi H, Orr HT (1993) Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. Nat Genet 5:254–258

David G, Abbas N, Stevanin G, Dürr A, Yvert G, Cancel G, Wever C, et al (1997) Cloning of the *SCA7* gene reveals a highly unstable CAG repeat expansion. Nat Genet 17:65–70

Deka R, Jin L, Shriver MD, Yu LM, DeCroo S, Hundrieser J, Bunker CH, et al (1995) Population genetics of dinucleotide $(dC\text{-}dA)_n\cdot(dG\text{-}dT)_n$ polymorphisms in world populations. Am J Hum Genet 56:461–474

Eichler EE, Holden JJA, Popovich BW, Reiss AL, Snow K, Thibodeau SN, Richards CS, et al (1994) Length of uninterrupted CGG repeats determines stability in the FMR1 gene. Nat Genet 8:88–94

Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361–372

Huntington's Disease Collaborative Research Group, The (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72:971–983

Ikeuchi T, Kazuhiro S, Takano H, Sasaki H, Tashiro K, Cancel G, Brice A, et al (1998) A novel long and unstable CAG/CTG trinucleotide repeat on chromosome 17q. Genomics 49:321–326

Kawaguchi Y, Okamota T, Taniwaki M, Aizawa M, Inoue M, Katayama S, Kawakami H, et al (1994) CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. Nat Genet 8:221–228

Kimmel M, Chakraborty R, Stivers DN, Deka R (1996) Dynamics of repeat polymorphisms under a forward-backward mutation model: within- and between-population variability at microsatellite loci. Genetics 143:549–555

Koide R, Ikeuchi T, Onodera O, Tanaka H, Kondo R, Ishikawa A, Hayashi T, et al (1994) Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). Nat Genet 6:9–13

Koob MD, Moseley ML, Schut LJ, Benzow KA, Bird TD, Day JW, Ranum LPW (1999) An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). Nat Genet 21:379–384

LaSpada AR, Wilson AM, Lubahn DB, Harding AE, Fishbeck KH (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature 352:77–79

Li CC (1976) First course in population genetics. Boxwood, Pacific Grove, CA, pp 1–15

Mahadevan M, Tsilfidis C, Sabourin L, Shutler G, Amemiya C, Jansen G, Neville C, et al (1992) Myotonic dystrophy mutation: an unstable CTG repeat in the 3′ untranslated region of the gene. Science 255:1253–1255

Monckton DG, Neumann R, Guram T, Fretwell N, Tamaki K, MacLeod A, Jeffreys AJ (1994) Minisatellite mutation rate variation associated with a flanking DNA sequence polymorphism. Nat Genet 8:162–170

Nakamoto M, Takebayashi H, Kawaguchi Y, Narumiya S, Taniwaki M, Nakamura Y, Ishikawa Y, et al (1997) A CAG/CTG expansion in the normal population. Nat Genet 17: 385–386

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89: 583–590

Orr HT, Chung M, Banfi S, Kwiatkowski TJ Jr, Servadio A, Beaudet AL, McCall AE, et al (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nat Genet 4:221–226

Pulst S-M, Nechiporuk A, Nechiporuk T, Gispert S, Chen X-N, Lopes-Cendes I, Pearlman S, et al (1996) Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. Nat Genet 14:269–276

Snedecor GW, Cochran WG (1976) Statistical methods, 5th ed. Iowa State University Press, Ames, pp 246–248

Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 8:1123–1128

Zhuchenko O, Bailevariancey J, Bonnen P, Ashizawa T, Stockton DW, Amos C, Dobyns WB, et al (1997) Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the a_{1A}-voltage-dependent calcium channel. Nat Genet 15:62–69