

Distribution and Evolution of CTG Repeats at the Myotonin Protein Kinase Gene in Human Populations

Ranjan Deka,^{1,11} Partha P. Majumder,² Mark D. Shriver,¹
David N. Stivers,³ Yixi Zhong,³ Ling M. Yu,¹ Ramiro Barrantes,⁴
Shih-Jiun Yin,⁵ Tetsuro Miki,⁶ Joachim Hundrieser,⁷
Clareann H. Bunker,⁸ Stephen T. McGarvey,⁹ Sameer Sakallah,¹⁰
Robert E. Ferrell,¹ and Ranajit Chakraborty³

¹Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania 15261; ²Anthropometry and Human Genetics Unit, Indian Statistical Institute, Calcutta, India; ³Human Genetics Center, University of Texas Health Science Center, Houston, Texas 77225; ⁴Instituto de Investigaciones en Salud, Universidad de Costa Rica, San Jose, Costa Rica; ⁵Department of Biochemistry, National Defense Medical Center, Taipei, Taiwan; ⁶Department of Geriatric Medicine, Osaka University Medical School, Osaka, Japan; ⁷Klinik für Abdominal und Transplantationschirurgie, Medizinische Hochschule, Hannover, Germany; ⁸Department of Epidemiology, University of Pittsburgh, Pittsburgh, Pennsylvania 15261; ⁹Program in Geographic Medicine, Brown University, Providence, Rhode Island 02906; ¹⁰Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania 15213

We have analyzed the CTG repeat length and the neighboring *Alu* insertion/deletion (+/-) polymorphism in DNA samples from 16 ethnically and geographically diverse human populations to understand the evolutionary dynamics of the myotonic dystrophy-associated CTG repeat. Our results show that the CTG repeat length is variable in human populations. Although the (CTG)₅ repeat is the most common allele in the majority of populations, this allele is absent among Costa Ricans and New Guinea highlanders. We have detected a (CTG)₄ repeat allele, the smallest CTG known allele, in an American Samoan individual. (CTG)_{≥19} alleles are the most frequent in Europeans followed by the populations of Asian origin and are absent or rare in Africans. To understand the evolution of CTG repeats, we have used haplotype data from the CTG repeat and *Alu*(+/-) locus. Our results are consistent with previous studies, which show that among individuals of Caucasian and Japanese origin, the association of the *Alu*(+) allele with CTG repeats of 5 and ≥19 is complete, whereas the *Alu*(-) allele is associated with (CTG)₁₁₋₁₆ repeats. However, these associations are not exclusive in non-Caucasian populations. Most significantly, we have detected the (CTG)₅ repeat allele on an *Alu*(-) background in several populations including Native Africans. As no (CTG)₅ repeat allele on an *Alu*(-) background was observed thus far, it was proposed that the *Alu*(-) allele arose on a (CTG)₁₁₋₁₃ background. Our data now suggest that the most parsimonious evolutionary model is (1) (CTG)₅-*Alu*(+) is the ancestral haplotype; (2) (CTG)₅-*Alu*(-) arose from a (CTG)₅-*Alu*(+) chromosome later in evolution; and (3) expansion of CTG alleles occurred from (CTG)₅ alleles on both *Alu*(+) and *Alu*(-) backgrounds.

Myotonic dystrophy (DM), one of the most common neuromuscular genetic disorders in adults, is characterized by myotonia accompanied by progressive weakness and wasting of distal muscles, cataracts, and cardiac arrhythmias (Harper

1989). It became the third example of human genetic disorders associated with trinucleotide repeat expansion when a normally polymorphic CTG repeat (5-37 repeats in normal individuals) located in the 3' untranslated region (UTR) of the myotonin protein kinase gene (DMPK) on 19q13.3 was found to be unstable and massively expanded (from 50 to several thousand repeats)

¹¹Corresponding author.
E-MAIL rdeka@helix.hgen.pitt.edu; FAX (412) 624-3020.

in DM patients (Brook et al. 1992; Buxton et al. 1992; Fu et al. 1992; Mahadevan et al. 1992). DM represents one of the most striking examples of anticipation (Harper et al. 1992). The question of meiotic drive at this locus, however, still remains an issue of controversy (Carey et al. 1994; Gennarelli et al. 1994; Hurst et al. 1995). Although the global incidence of DM is not well known, it varies from ~1/8000 among Caucasians (Harper 1989) to ~1/18000 among Japanese (Davies et al. 1992). Among Africans, DM is extremely rare; to date only one family from Nigeria with DM has been reported (Dada 1973). An *Alu* insertion/deletion (+/-) polymorphism, intragenic to the DMPK gene and located 5 kb telomeric to the CTG repeat, was shown to be in complete linkage disequilibrium with the DM mutation (Mahadevan et al. 1993). A single origin of the DM mutation was postulated based on an initial observation that all disease chromosomes in Eurasian patients were found on the *Alu*(+) background (Harley et al. 1992; Yamagata et al. 1992; Lavedan et al. 1994). Imbert et al. (1993) further examined the issue of the origin of the DM mutation by analyzing linkage disequilibrium between normal CTG repeat alleles, the *Alu*(+/-) polymorphism, and a CA repeat polymorphism 90 kb distal to the CTG repeat. The smallest CTG repeat allele (repeat 5) and the large normal alleles (repeats 16–36) were found to be completely associated with the *Alu*(+) allele, whereas the intermediate alleles (11–13 repeats) were more often (99%) associated with the *Alu*(-) allele. On the basis of this observation, Imbert et al. (1993) proposed that through a few early events, the 5-repeat allele mutated to an allele of 19 to 30 repeats, and in turn, this later group of alleles has served as precursor of the expanded disease alleles by recurrent mutations. The *Alu*(+/-) polymorphism probably evolved through the deletion of an *Alu* element early in human evolution, because the *Alu*(+) allele is the only allele present at the orthologous location in nonhuman primates, and this allele is present in most human populations (Rubinsztein et al. 1994). Recent studies in global populations show that the non-random association between CTG repeats and the *Alu*(+/-) polymorphism is not always observed (Rubinsztein et al. 1994; Goldman et al. 1995; Zerylnick et al. 1995). Most strikingly, in the lone reported DM family from Nigeria, the expanded alleles were found on an *Alu*(-) background (Krahe et al. 1995). These findings cast doubt on the validity of the hypothesis proposed

by Imbert et al. (1993) and the notion of a particular haplotype predisposed to expansion (Neville et al. 1994). To better understand the evolution and dynamics of the CTG repeats at the DMPK gene, we have analyzed the CTG repeat and the *Alu*(+/-) polymorphism in DNA samples from ~800 individuals belonging to 16 geographically and ethnically diverse populations encompassing all major racial groups.

RESULTS

CTG Repeat Distribution in Human Populations

The CTG repeat length was determined in a total of 802 individuals belonging to five major groups of human populations, namely Caucasians, Asian Mongoloids, Africans, American Indians, and Pacific islanders. Twenty-seven alleles in the range of 4–34 repeats were detected in the pooled sample. The population-specific allele frequencies, grouped in six repeat classes, are given in Table 1. The majority of the populations show two major peaks, one at the smallest allele, 5 repeats, and the other in the class of 11–14 repeats. The Caucasians, excepting the Indian Brahmans, show a third mode at the 19–30 repeat class. However, there are a few exceptions. The Costa Rican Cabecar population has alleles exclusively in the 11–14 repeat class; allele 5 is absent among the New Guinea highlanders and is present at a relatively low frequency (6.5%) among the Pehuenche Indians. These two populations are also characterized by high frequencies of alleles of 11–14 repeats, 76% and 81%, respectively. Another interesting finding is a four-repeat allele in an American Samoan individual. With the exception of two populations, Brazilian white ($\chi^2 = 3.77$; $P = 0.04$) and New Guinea highlander ($\chi^2 = 7.68$; $P < 0.001$), no significant deviation from Hardy-Weinberg expectations was observed in any population.

Because of their being implicated as the alleles predisposed to expansion (Imbert et al. 1993), the distribution of large normal alleles in the class ≥ 19 repeats is of interest. When the frequencies of this class of alleles are compared among the populations for whom the incidence of DM is known, the observed trend is in accordance with the expected. Globally, western Europeans are known to have the highest incidence of DM followed by the Japanese. The incidence of DM is lowest among Africans. This is parallel to

Table 1. Frequency of CTG Repeat Alleles in 16 Human Populations

Populations, their anthropological affinity (no. of chromosomes)	Range of repeat size		No. of alleles	Repeat size class and their frequencies					
	min	max		≤5	6–10	11–14	15–18	19–30	≥31
Caucasians									
CEPH (158)	5	28	14	.399	.038	.430	.057	.076	—
German (104)	5	27	15	.327	—	.442	.019	.212	—
Brazilian white (100)	5	25	15	.390	.050	.460	.030	.070	—
Indian Brahman (72)	5	27	12	.333	.042	.472	.111	.042	—
Asian Mongoloids									
Chinese (102)	5	23	12	.294	.039	.569	.088	.010	—
Japanese (176)	5	28	15	.176	.017	.656	.074	.074	—
Kachari (86)	5	26	10	.302	.128	.523	.023	.023	—
Africans									
Benin (100)	5	18	11	.320	.140	.520	.020	—	—
Sokoto (106)	5	19	11	.264	.198	.500	.028	.009	—
Brazilian black (84)	5	28	13	.191	.083	.655	.012	.060	—
American Indians									
Dogrib Indian (100)	5	34	9	.310	.020	.620	—	.020	.030
Cebecar (104)	11	13	3	—	—	1.00	—	—	—
Pehuenche Indian (92)	5	27	9	.065	.087	.761	.022	.065	—
Pacific islanders									
New Guinea highlander (100)	9	19	10	—	.130	.810	.040	.020	—
American Samoa (98)	4	20	9	.327	.031	.449	.184	.010	—
Western Samoa (100)	5	25	8	.350	.040	.520	.080	.010	—

the observed frequencies of ≥ 19 repeat alleles in our examined populations, from a frequency of 21% among the Germans to 7.4% among the Jap-

anese to virtual absence among Native Africans. The relatively higher frequency (6%) of large alleles in the Brazilian black population can be attributed to Caucasian admixture. Although ethnically closer to the Japanese, the absence of these large alleles among Chinese is a notable observation because DM is reported to be very rare in China (Ashizawa and Epstein 1991). Despite significant intrapopulation variations among the major population groups, in general, the frequencies of ≥ 19 repeat alleles are significantly elevated in Caucasians as compared with other major ethnic groups. The prevalence of DM among American Indians and Pacific islanders is not known. Nevertheless, one of these populations, the Dogrib Indians,

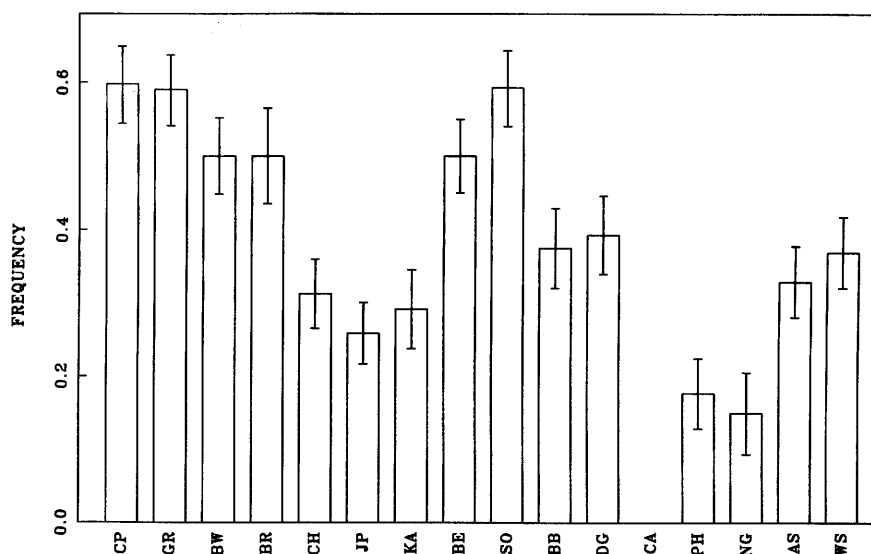


Figure 1 Frequency distribution (± 1 s.d.) of the *Alu(+)* allele at the DMPK locus in 16 human populations.

Nevertheless, one of these populations, the Dogrib Indians,

Table 2. CTG-Repeat and *Alu*(+/-) Haplotype Frequencies in Four Caucasian Populations

(CTG) _n ^a	Populations ^b														
	CP (82)				GR (100)				BW (94)				BR (60)		
	(+)	(-)	<i>D/D</i> _{max}	(+)	(-)	<i>D/D</i> _{max}	(+)	(-)	<i>D/D</i> _{max}	(+)	(-)	<i>D/D</i> _{max}	(+)	(-)	<i>D/D</i> _{max}
5	.390	—	1.00 ^c	.320	—	1.00 ^c	.404	—	1.00 ^d	.277	.023	0.85 ^c	.277	.023	0.85 ^c
6	—	—	—	—	—	—	.010	—	1.00	—	.017	-1.00	—	.017	-1.00
7	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
8	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
10	.037	—	1.00	—	—	—	—	—	—	—	0.17	-1.00	—	0.17	-1.00
11	.001	.134	-0.99 ^d	—	.100	-1.00 ^c	—	.106	-1.00 ^c	—	.217	-1.00 ^c	—	.217	-1.00 ^c
12	—	.146	-1.00 ^c	—	.130	-1.00 ^c	—	.096	-1.00 ^c	.010	.073	-0.78	—	.073	-0.78
13	.012	.110	-0.84 ^c	—	.150	-1.00 ^c	.021	.202	-0.81 ^c	.042	.125	-0.49	—	.125	-0.49
14	.049	.012	-0.50	.040	.030	-0.03	—	.021	-1.00	.018	.015	-0.08	—	.015	-0.08
15	—	—	—	—	—	—	.010	.021	-0.33	.033	—	1.00	—	.033	1.00
16	.037	—	1.00	.010	—	1.00	—	—	—	.036	.014	0.45	—	.014	0.45
17	—	—	—	—	—	—	—	—	—	.033	—	1.00	—	.033	1.00
19	—	—	—	.030	—	—	—	.011	-1.00	—	—	—	—	—	—
20	.012	—	1.00	.020	—	1.00	.011	—	1.00	—	—	—	—	—	—
21	.024	—	1.00	.060	—	1.00 ^d	.011	—	1.00	—	—	—	—	—	—
22	.024	—	1.00	.060	—	1.00 ^d	.011	—	1.00	—	—	—	—	—	—
23	—	—	—	.010	—	1.00	.011	—	1.00	.033	—	1.00	—	.033	1.00
24	.012	—	1.00	.010	—	1.00	—	—	—	—	—	—	—	—	—
25	—	—	—	—	—	—	.011	—	1.00	—	—	—	—	—	—
26	—	—	—	.010	—	1.00	—	—	—	—	—	—	—	—	—
27	—	—	—	.020	—	1.00	—	—	—	.017	—	—	—	.017	—

^a(CTG)_n indicates number of CTG repeats.

^b(CP) CEPH parents; (GR) German; (BW) Brazilian white; (BR) Indian Brahman. Numbers in parentheses indicate number of chromosomes studied. (+) indicates *Alu*(+); (-) indicates *Alu*(-). *D/D*_{max} values are always in reference to the (CTG) repeat alleles in association with the *Alu*(+) allele.

^c χ^2 for *D* is significant at *P* < 0.01.

^d χ^2 for *D* is significant at *P* < 0.05.

has large alleles (repeats 32, 33, 34) that have not been observed in other populations. The Pehuenche Indians also have a comparably higher frequency (6.5%) of ≥ 19 repeat alleles. It would be interesting to screen these populations for the presence of DM.

Association between CTG Repeats and *Alu*(+/-) in Global Populations

The *Alu*(+/-) polymorphism was determined in a subset of the individuals examined (677 of 802). Frequencies of the insertion allele in the populations studied are shown in Figure 1. In general, the proportion of the insertion allele is almost equal or slightly higher than the deletion allele among the Caucasians and Africans, with the exception of Brazilian blacks, who have a significantly lower frequency of the insertion allele. It is known that Brazilian blacks are admixed with Caucasians; it is plausible that both of these populations are admixed with the local American Indians as well. In contrast, the Asiatic, American

Indian, and Pacific island populations are characterized by a significantly higher frequency of the deletion allele, with the Cabecar being monomorphic for this allele. All populations are in Hardy-Weinberg equilibrium.

We have determined haplotype frequencies of the CTG repeats and the *Alu*(+/-) polymorphism using the algorithm of Long et al. (1995) of haplotype frequency estimation from two-locus genotype data. Associations of the CTG repeat alleles with the *Alu*(+) alleles were measured by the relative value of linkage disequilibria (D/D_{\max}) coefficients (Lewontin 1964). The haplotype frequencies along with the D/D_{\max} values are presented in Tables 2-6. The results can be summarized as follows: (1) Although the (CTG)₅ repeat is significantly associated with the *Alu*(+) allele in all populations, this repeat is also found with the *Alu*(-) allele in several populations, including the Indian Brahman, Benin African, Brazilian black, and Pehuenche Indian samples. Thus, the association of the (CTG)₅ repeat allele with *Alu*(+) is not exclusive for all human popu-

Table 3. CTG Repeat and *Alu*(+/-) Haplotype Frequencies in Three Mongoloid Populations

(CTG) _n	Populations ^a								
	CH (96)			JP (104)			KA (72)		
	(+)	(-)	D/D_{\max}	(+)	(-)	D/D_{\max}	(+)	(-)	D/D_{\max}
5	.292	—	1.00 ^b	.139	—	1.00 ^b	.264	—	1.00 ^b
7	—	.021	-1.00	—	—	—	—	—	—
8	—	.010	-1.00	—	—	—	—	—	—
10	—	.010	-1.00	.009	.009	0.34	—	.125	-1.00 ^b
11	—	.073	-1.00	.019	.195	-0.66 ^c	—	.181	-1.00 ^c
12	—	.260	-1.00 ^b	—	.296	-1.00 ^b	—	.222	-1.00 ^b
13	—	.156	-1.00 ^b	—	.213	-1.00 ^b	—	.097	-1.00
14	.010	.083	-0.64	.009	—	1.00	—	.069	-1.00
15	—	.052	-1.00	—	.019	-1.00	—	—	—
16	—	.011	-1.00	.009	.009	0.34	—	.014	-1.00
17	—	—	—	.009	—	1.00	0.14	—	1.00
18	.011	.011	0.28	—	—	—	—	—	—
21	—	—	—	.009	—	1.00	—	—	—
22	—	—	—	.009	—	1.00	—	—	—
23	—	—	—	.019	—	1.00 ^c	—	—	—
24	—	—	—	—	—	—	.014	—	1.00
26	—	—	—	.019	—	1.00 ^c	—	—	—
27	—	—	—	.009	—	1.00	—	—	—

See Table 2 for additional explanation.
^a(CH) Chinese; (JP) Japanese; (KA) Kachari.
^b χ^2 for D is significant at $P < 0.01$.
^c χ^2 for D is significant at $P < 0.05$.

Table 4. CTG Repeat and *Alu*(+/-) Haplotype Frequencies in Three Populations of African Origin

(CTG) _n	Populations ^a								
	BE (100)			SO (86)			BB (80)		
	(+)	(-)	<i>D/D</i> _{max}	(+)	(-)	<i>D/D</i> _{max}	(+)	(-)	<i>D/D</i> _{max}
5	.268	.052	.068 ^b	.256	—	1.00 ^b	.162	.012	.088 ^b
6	—	.020	-1.00	—	—	—	—	—	—
7	.009	.011	-0.10	.023	—	1.00	—	.025	-1.00
8	.011	.019	-0.28	.023	—	1.00	.038	—	1.00 ^c
9	—	—	—	—	.023	-1.00	—	—	—
10	.057	.013	0.62	.114	.026	0.55	.013	.013	0.20
11	.069	.081	-0.08	.054	.121	-0.48 ^b	—	.125	-1.00 ^b
12	.013	.147	-0.83 ^b	.012	.162	-0.89 ^b	.025	.162	-0.64 ^c
13	.032	.058	-0.29	.030	.052	-0.39	—	.238	-1.00 ^b
14	.031	.089	-0.49	.070	—	1.00 ^c	.079	.034	0.52 ^c
15	—	—	—	.011	.012	-0.16	—	—	—
16	—	.010	-1.00	—	—	—	—	.012	-1.00
18	.010	—	1.00	—	—	—	—	—	—
19	—	—	—	—	.011	-1.00	—	—	—
21	—	—	—	—	—	—	.021	.004	0.75
23	—	—	—	—	—	—	.013	—	1.00
24	—	—	—	—	—	—	.012	—	1.00
28	—	—	—	—	—	—	.012	—	1.00

See Table 2 for additional explanation.
^a(BE) Benin; (SO) Sokoto; (BB) Brazilian black.
^b χ^2 for *D* is significant at *P* < 0.01.
^c χ^2 for *D* is significant at *P* < 0.05.

lations. (2) In most populations $\geq 90\%$ of the alleles in the (CTG)₁₁₋₁₃ repeat class are associated with the *Alu*(-) allele. Among Native Africans, however, this association is considerably weaker; 29% and 22% of (CTG)₁₁₋₁₃ chromosomes are present on the *Alu*(+) background among Nigerians of Benin and Sokoto, respectively. (3) (CTG) _{≥ 19} repeats among the Caucasian, Asian Mongoloid, and Pacific island populations are present on the *Alu*(+) background, with a single exception. One Brazilian white (CTG)₁₉ repeat allele was found in association with the *Alu*(-) allele. CTG repeats in the ≥ 19 repeat class are virtually absent in the two Native African populations. The single (CTG)₁₉ repeat found in the Sokoto population from Nigeria was associated with the *Alu*(-) allele. In contrast, the Brazilian black population has a significantly higher frequency of ≥ 19 repeats, which are exclusively associated with the *Alu*(+) allele. Interestingly, as mentioned above, in two American Indian populations, the Dogrib and Pehuenche, larger CTG

repeat alleles in the ≥ 19 class are present at comparatively higher frequencies. Among the Dogrib, all of these alleles are exclusively associated with the *Alu*(+) allele. However, a pronounced disassociation was observed among the Pehuenche, among whom 41% of these larger CTG repeats are associated with the *Alu*(-) allele. (4) The other CTG repeat arrays show a largely random distribution between *Alu*(+/-) alleles. (5) Linkage disequilibrium values between the (CTG)_n and *Alu*(+/-) loci are highly significant [*P* < 0.002, by the likelihood ratio test of over all linkage disequilibrium (Long et al. 1995)] in all populations studied (data not shown).

DISCUSSION

There are two principal objectives of this study. First, we report frequencies of CTG repeats at the DMPK locus in normal individuals from population samples of diverse ethnic and geographic origin. Second, our analysis of CTG repeat alleles in

Table 5. CTG Repeat and *Alu*(+/-) Haplotype Frequencies in Three American Indian Populations

(CTG) _n	Populations ^a								
	DG (84)			CA (96)			PH (62)		
	(+)	(-)	<i>D/D</i> _{max}	(+)	(-)	<i>D/D</i> _{max}	(+)	(-)	<i>D/D</i> _{max}
5	.321	—	1.00 ^b	—	—	—	.064	.017	0.75 ^b
10	—	.024	-1.00	—	—	—	—	.065	-1.00
11	—	.202	-1.00 ^b	—	.042	-1.00	.033	.402	-0.57
12	.012	.167	-0.83 ^b	—	.125	-1.00	—	.032	-1.00
13	—	.214	-1.00 ^b	—	.833	-1.00	.001	.274	-1.00 ^c
14	—	—	—	—	—	—	.016	—	1.00 ^c
16	—	—	—	—	—	—	.016	—	1.00 ^c
24	—	—	—	—	—	—	.047	.017	0.68 ^b
27	.023	—	1.00	—	—	—	—	.016	-1.00
32	.013	—	1.00	—	—	—	—	—	—
33	.012	—	1.00	—	—	—	—	—	—
34	.012	—	—	—	—	—	—	—	—

See Table 2 for additional explanation.
^a(DG) Dogrib; (CA) Cabecar; (PH) Pehuenche Indian.
^b χ^2 for *D* is significant at $P < 0.01$.
^c χ^2 for *D* is significant at $P < 0.05$.

association with the intragenic *Alu* (+/-) polymorphism in the examined populations provide insight into the evolutionary origin of CTG repeat arrays. Such information is helpful in understanding the molecular epidemiology and natural history of DM.

With regard to DM, available data show a positive correlation between the incidence of the disease and the frequency of large, normal CTG repeats (size ≥ 19) in European and Japanese populations (Davies et al 1992; Imbert et al. 1993; Yamagata et al. 1996). Thus, it was postulated that normal alleles of size ≥ 19 CTG repeats constitute a reservoir for recurrent expansion mutations, which compensate for the loss of the disease alleles attributable to the negative effect of anticipation, and consequently maintain the disease incidence at a constant rate (Imbert et al. 1993; Neville et al. 1994). These observations were based on the results of the initial studies in populations with known incidence of DM, which were conducted soon after the discovery that expanded CTG repeats cause DM. The absence or rarity of DM among Africans associated with a significant paucity of the large normal alleles in these populations confirmed this proposition (Goldman et al. 1994; Watkins et al. 1995). The results of subsequent studies in other populations

are in agreement with this hypothesis (Rubinsztein et al. 1994; Zerylnick et al. 1995). In this study we have examined a wider range of human populations of defined origin. With the exception of Europeans, Japanese, and Native Africans, the incidence of DM in the other populations is not well documented. A questionnaire survey reported that DM is more prevalent among the North Eurasians and rare in African and Southeast Asian populations (Ashizawa and Epstein 1991). As such, if frequency of (CTG) _{≥ 19} alleles is an indicator of prevalence of DM, our study of the distribution of CTG repeats provides an indirect source of information on population-specific prevalence of this disease.

The results of our study show that although there is considerable interpopulation variation in the frequencies of CTG repeats, (CTG)₅ is the smallest and most common allele in the majority of the populations, and the combined frequency of the (CTG)₁₁₋₁₄ repeats is the most prevalent of all classes of CTG repeat arrays in all populations, irrespective of their origin. These findings are in agreement with the results of previous investigations. However, there are a few exceptions. The Cabecar of Costa Rica are unimodal for the (CTG)₁₁₋₁₄ repeat class; (CTG)₅ is absent among the New Guinea highlanders, and this repeat has

Table 6. CTG Repeat and *Alu*(+/-) Haplotype Frequencies in Three Pacific Islander Populations

(CTG) _n	Populations ^a								
	NG (40)			AS (94)			WS (100)		
	(+)	(-)	<i>D/D</i> _{max}	(+)	(-)	<i>D/D</i> _{max}	(+)	(-)	<i>D/D</i> _{max}
4	—	—	—	.011	—	1.00	—	—	—
5	—	—	—	.309	—	1.00 ^b	.350	—	1.00 ^b
9	—	.075	-1.00	—	.032	-1.00	—	.040	-1.00
10	—	.050	-1.00	—	—	—	—	—	—
11	.050	.250	0.05	—	.043	-1.00	—	.090	-1.00 ^c
12	—	.150	-1.00	—	.021	-1.00	.010	.080	-0.70
13	—	.100	-1.00	—	.287	-1.00 ^b	—	.280	-1.00 ^b
14	—	.200	-1.00	—	.096	-1.00 ^c	—	.060	-1.00
15	.025	—	1.00 ^c	—	.191	-1.00 ^b	—	.080	-1.00 ^c
17	.050	.025	0.61 ^b	—	—	—	—	—	—
19	.025	—	1.00 ^c	—	—	—	—	—	—
20	—	—	—	.010	—	1.00	—	—	—
25	—	—	—	—	—	—	.010	—	1.00

See Table 2 for additional explanation.
^a(NG) New Guinea highlanders; (AS) American Samoa; (WS) Western Samoa.
^b χ^2 for *D* is significant at $P < 0.01$.
^c χ^2 for *D* is significant at $P < 0.05$.

a particularly low frequency (6.5%) among the Pehuenche Indians. The most likely explanation for these observed departures from the norm is the effect of genetic drift on these populations, whose effective population sizes are much smaller. Interestingly, among the Japanese as well, (CTG)₁₂ is the most frequent repeat (29%) followed by (CTG)₅. Zerylnick et al. (1995) also have reported a lower frequency of (CTG)₅ among Native Americans and Tibetans. However, their remark that (CTG)₅ may never have been brought to the Americas by the Amerind or Na-Dene migrations is not corroborated by our results. For instance, the Dogrib, a Na-Dene population in our study, has a high frequency (31%) of the (CTG)₅ allele. Notwithstanding these few exceptions, ours as well as the previous studies show that in general, (CTG)₅ is the most common allele in human populations. This observation, further substantiated by conservation of a similar size repeat in the homologous DMPK gene in mouse (Jansen et al. 1992), may suggest that (CTG)₅ is the progenitor of all alleles. Interestingly, we have detected a novel (CTG)₄ repeat allele in an American Samoan individual. This allele could have arisen by slippage from a five-repeat allele.

The spectrum of the distribution of (CTG)_{≥19} repeats in the populations with known incidence of DM supports the argument that a comparison of the frequency of these repeats may serve as an indicator of the incidence of the disease in various populations. Thus, for populations among which the incidence of DM has not yet been documented, our results predict a lower prevalence of DM among the Costa Rican Cabecar and Pacific islanders. On the other hand, relatively higher incidence of CTG repeats in the range ≥24 among the Dogrib and Pehuenche Indians may be indicative of presence of DM.

Our second objective addresses the issue of the evolution of CTG repeats at the DM locus using haplotype frequencies at the linked CTG repeat and the *Alu*(+/-) loci. Although a number of similar studies have been published previously (Imbert et al. 1993; Neville et al. 1994; Rubinsztein et al. 1994; Goldman et al. 1995; Zerylnick et al. 1995; Yamagata et al. 1996), several of these were limited to single populations, and in most cases haplotypes were constructed using homozygotes at one of the loci, excluding double heterozygotes. Haplotype frequencies estimated from such truncated data obviously do not represent their true frequencies in populations.

Therefore, the understanding of the evolutionary dynamics of the CTG repeats has remained incomplete. For the first time, our study reports the estimated haplotype frequencies in healthy subjects on a worldwide scale; the 16 studied populations belonging to five major racial groups encompass a broad spectrum of human genome diversity. Our findings are consistent with previous studies (Imbert et al. 1993; Rubinsztein et al. 1994; Yamagata et al. 1996) that in general, among individuals of Caucasian and Japanese ancestry, the association of the *Alu(+)* allele with CTG repeats of 5 and ≥ 19 is complete (as indicated by $D/D_{\max} = 1.00$), whereas the *Alu(-)* allele is restricted to the CTG repeat class 11–16. Similarly, our results are in agreement with the findings (Rubinsztein et al. 1994; Goldman et al. 1995; Zerylnick et al. 1995) that in non-Caucasian populations, particularly among Africans, the nonrandom associations of (1) *Alu(-)* and (CTG)_{11–13} repeats, and (2) *Alu(+)* and (CTG) _{≥ 19} repeats are not exclusive. This is evident from the values of D/D_{\max} between CTG-repeat alleles and the *Alu(+)* allele as shown in Tables 2–6. In addition to these, for the first time we have observed the presence of the (CTG)₅–*Alu(-)* haplotype in several of the populations, such as Indian Brahman (Caucasian), Benin, and Brazilian black (African) and Pehuenche Indian (American Indian).

The finding of (CTG)₅ in the *Alu(-)* background is significant because it provides a new dimension to the evolutionary history of CTG repeats. It refutes the model of Imbert et al. (1993), which proposed that (1) CTG repeats evolved from two founder chromosomes, (CTG)₅ in the *Alu(+)* background and (CTG)_{11–13} in the *Alu(-)* background, and (2) (CTG) _{≥ 19} repeats evolved from (CTG)₅ repeats because they share similar haplotypes. After the finding of (CTG)_{11–13} repeats in association with the *Alu(+)* allele and (CTG) _{≥ 19} alleles in association with the *Alu(-)* allele, a second model was developed by Rubinsztein et al. (1994), which proposed that (1) (CTG)_{11–13} repeats evolved from the (CTG)₅ repeats, (2) *Alu(-)* allele arose on a (CTG)_{11–13} allele, and (3) (CTG) _{≥ 9} alleles evolved from (CTG)_{11–13} repeats in both *Alu* insertion and deletion backgrounds. Despite their differences, both models have assumed that the (CTG)₅–*Alu(-)* haplotype is not present in human populations. Our finding leads us to propose a somewhat different model. We propose that the most parsimonious evolutionary model is (1) (CTG)₅–*Alu(+)* is the ances-

tral chromosome because it is the most common haplotype in most human populations and, among nonhuman primates only the *Alu(+)* allele is observed (Rubinsztein et al. 1994), (2) (CTG)₅–*Alu(-)* arose from a (CTG)₅–*Alu(+)* chromosome later in evolution, and (3) CTG alleles independently expanded from (CTG)₅ alleles on both *Alu(+)* and *Alu(-)* backgrounds, although the rates of expansion may have been different as is evidenced by the relative magnitude of the linkage disequilibrium coefficients given in Tables 2–6.

We find that the hypothesis of Imbert et al. (1993) on the evolution of (CTG) _{≥ 19} repeats from (CTG)₅ through a few initial events is not tenable for the following reasons. There is a continuous distribution of CTG repeats along both *Alu(+)* and *Alu(-)* lineages, especially in the African populations. Second, the fitting of mutation drift models following the approach of Shriver et al. (1993) show that with the exception of two Caucasian populations, the unrelated Centre d'Etudes du Polymorphisme Humain (CEPH) parents and the Brazilian white (cosmopolitan and admixed populations), in all of the populations the observed numbers of CTG alleles agree with the expectations derived under a single step stepwise mutation model (data not shown). This result is at variance with large multistep jumps as proposed by Imbert et al. The lack of correspondence of (CTG)_{<6}–*Alu(+)* and (CTG)_{>18}–*Alu(+)* haplotype frequencies in the 16 populations studied here (Fig. 2) is also discordant with the hypothesis of Imbert et al. If (CTG)_{>18}–alleles evolved from (CTG)_{<6} alleles on the *Alu*-insertion background, we should expect a positive correlation of (CTG)_{<6}–*Alu(+)* and (CTG)_{>18}–*Alu(+)* haplotype frequencies across populations. In contrast, under parallel expansions of CTG repeats in the two (CTG)_{<6}–*Alu(+)* and (CTG)_{<6}–*Alu(-)* lineages, no such correlation of (CTG)_{<6}–*Alu(+)* and (CTG)_{>18}–*Alu(+)* haplotype frequencies across populations is expected. As shown in Figure 2, this is in agreement with our data on 16 populations where the rank correlation (Kendall's p) of these two haplotype frequencies across 16 populations is 0.35 ($P > 0.11$). Our model of CTG repeat evolution is more similar to the one proposed by Rubinsztein et al. (1994). The difference perhaps has resulted from the absence of (CTG)₅–*Alu(-)* haplotype in the population samples examined by Rubinsztein et al.

Our proposed parsimonious model of evolution of CTG repeat alleles may be criticized on

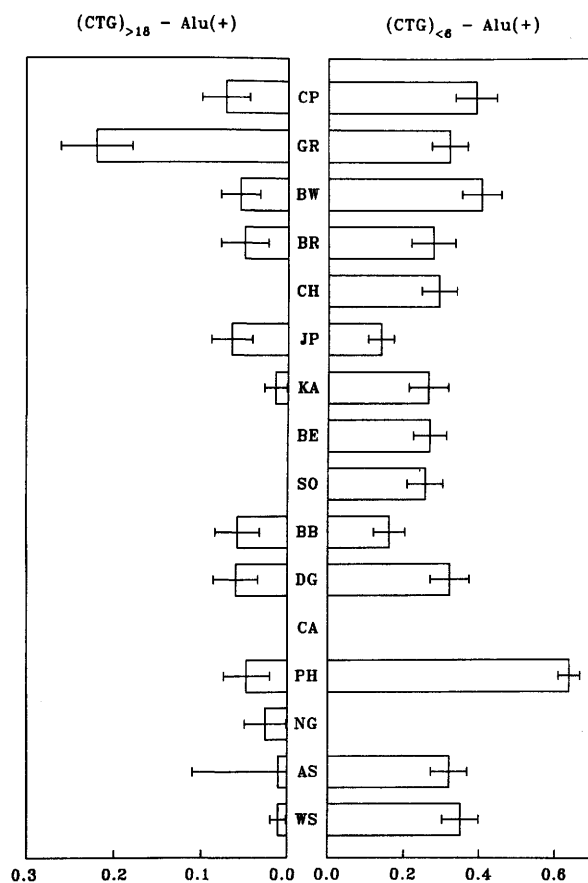


Figure 2 Frequency distribution (± 1 s.d.) of $(CTG)_{>18}$ and $(CTG)_{<6}$ alleles on $Alu(+)$ background in 16 human populations.

the ground that we have de-emphasized the possibility of recombination in this gene region. There are several lines of evidence suggesting that recombination is unlikely to explain the data presented here. First, as mentioned before, it is quite evident that the $(CTG)_5$ repeat is the oldest allele at the repeat region of the DMPK gene, and $Alu(+)$ is the ancestral allele at the $Alu(+/-)$ polymorphic site, making the $(CTG)_5-Alu(+)$ haplotype the most likely ancestral haplotype during human evolution. Therefore, a mutation causing the deletion of an Alu element, occurring on the $(CTG)_5-Alu(+)$ background, is a likely scenario that would explain the data presented here. The observation that the $(CTG)_5-Alu(+)$ is the most common haplotype in 11 of the 16 populations examined here supports this assertion, as the oldest haplotype also has the highest probability of being the most frequent (see Tables 2–6). Second, the possibility that the $Alu(-)$ mutation occurred in the background to $(CTG)_{11-13}$ alleles (Rubinsz-

tein et al. 1994), and subsequent recombinations have created the $(CTG)_5-Alu(-)$ haplotype, is also not supported by our data. Under this scenario, the progenitor haplotype $(CTG)_{11-13}-Alu(+)$, on which such a mutation occurred, should be found more globally [especially in non-African populations, as the $Alu(-)$ allele is present everywhere] compared to the ones observed in our study and in that of Goldman et al. (1995). Third, the Alu site is intragenic to the DMPK gene and is located 5 kb telomeric to the CTG repeat site (Mahadevan et al. 1993). In this short distance, the recombination rate is small and there is no suggestion of a recombinational hotspot within this gene region.

The distribution of CTG repeats at the DMPK locus also reflects the demographic histories of populations. In general, we have observed a reduced degree of variability among American Indians and Pacific islanders, whose effective population sizes are smaller, compared to large and cosmopolitan populations such as the Africans, Caucasians, and Asian Mongoloids. This can be demonstrated through computations of gene diversity values at the CTG repeat site and the $Alu(+/-)$ site and haplotype diversity (unbiased estimates; Nei 1978), using the frequencies shown in Tables 2–6. Table 7 shows these computations, along with the standard errors of these estimates and the number of haplotypes observed in the 16 populations examined here. Even though the differences of the gene diversity and haplotype diversities across several of these populations are not significant, some trends are instructive. For example, the populations of African ancestry have the largest haplotype diversity accompanied with larger number of observed haplotypes. This is consistent with the hypothesis of an African origin of the human species. The smallest level of diversity in the Cabecar population is particularly noteworthy. We contend that this reflects an effect of genetic drift in this isolated population, as reduced gene diversity is observed in this population at several additional microsatellite loci (Deka et al. 1995a). That it is not a sampling artifact is evident from the analysis of triplet repeat polymorphisms at the HD and SCA loci, where the Cabecars exhibit levels of gene diversity and numbers of alleles larger than those observed in other small populations (R. Deka, M.D. Shriver, R. Barrantes, and R. Chakraborty, unpubl.). The absence of the $Alu(+)$ allele, together with the small number of haplotypes in the Cabecar, are probably effects of ge-

Table 7. Estimates of Gene and Haplotype Diversity and Number of Haplotypes in 16 Populations

Populations	Percent gene diversity (s.e.)		No. of haplotypes	Percent haplotype diversity (s.e.)
	CTG repeat	<i>Alu</i> (+/-)		
CEPH	79.5 (3.3)	48.7 (2.3)	14	79.9 (3.4)
German	84.2 (2.3)	48.9 (1.9)	15	84.5 (2.4)
Brazilian white	77.1 (3.2)	50.5 (0.8)	17	78.0 (3.3)
Indian Brahman	83.4 (2.8)	50.9 (1.2)	17	86.1 (2.8)
Chinese	81.3 (2.1)	43.5 (3.6)	13	81.5 (2.1)
Japanese	80.8 (1.8)	38.8 (4.2)	17	81.5 (1.9)
Kachari	82.9 (2.1)	41.9 (4.6)	9	82.9 (2.1)
Benin	82.8 (2.1)	50.5 (0.7)	19	88.3 (1.9)
Sokoto	85.0 (1.7)	48.8 (2.1)	16	87.7 (1.9)
Brazilian black	85.6 (1.7)	47.5 (2.8)	18	87.5 (1.8)
Dogrib Indian	78.6 (2.0)	48.3 (2.4)	10	79.0 (2.0)
Cabecar	29.2 (5.6)	0.0 (0.0)	3	29.2 (5.6)
Pehuenche Indian	73.0 (4.1)	29.6 (6.3)	13	76.2 (4.0)
New Guinea highlander	84.4 (3.3)	26.2 (8.1)	11	87.2 (2.8)
American Samoa	78.1 (2.1)	44.7 (3.4)	9	78.1 (2.1)
Western Samoa	77.9 (2.5)	47.1 (2.6)	9	78.1 (2.5)

netic drift in this isolated small population. The Dogrib of Northwestern Canada exhibit a substantial haplotype diversity (79%), a high frequency of the (CTG)₅-*Alu*(+) haplotype (32%), and they also have several larger (CTG) repeat alleles (27 repeats or larger). This is probably caused by Caucasian admixture in this population, for which some evidence exists (Szathmary et al. 1983). Our study further supports an interpretation provided by Goldman et al. (1995) on the absence of (CTG)₁₁₋₁₃-*Alu*(+) and (CTG)_{≥19}-*Alu*(-) haplotypes among non-African populations, which asserts that the ancestors of the present-day non-African populations left Africa with a limited number of African chromosomes at the time of the first split of the African and non-African populations. Furthermore, if (CTG)₅-*Alu*(-) arose from (CTG)₅-*Alu*(+) background, then our data also provide evidence that the *Alu*(-) mutation took place in African populations. Migration from "out of Africa" possibly took place after a few steps of CTG expansion along the *Alu*(-) lineage. Furthermore, our data support the argument in favor of a multiple origin of the DM mutation (Krahe et al. 1995). We have stated previously that the DM mutation among Europeans and Japanese had a common Eurasian origin (Yamagata et al. 1996), whereas the African mutation represents an independent mutational event that arose on a separate haplo-

typic background (Krahe et al. 1995). On the basis of a significantly higher prevalence of DM among Europeans and Japanese compared to its rarity among Africans, it is likely that the Eurasian mutation is more ancient than the African mutation.

METHODS

Population Samples

The 16 human populations studied belong to five major groups, namely Caucasian, Mongoloid, African, American Indian, and Pacific islander. The Caucasians are represented by unrelated parents of the CEPH cohort, a German sample from northern Germany, a white population from Brazil, and a Brahman sample from northern India. The Mongoloid populations are a Chinese sample of Han origin, a Japanese sample from around Osaka, Japan, and the Kachari, a tribal populations from northeast India. The Africans are represented by two west African linguistically diverse populations, the Benin and the Sokoto from Nigeria, and a black sample from Brazil. The American Indian samples consist of the Dogrib Indian drawn from the Northwest Territories, Canada, who represents the Na-Dene group, the Pehuenche Indian from Chile, who constitute a branch of Araucanian Indians, and the Cabecar, a tribal population from Costa Rica. The Pacific Islands are represented by two Samoan groups drawn from villages distributed throughout American Samoa and Western Samoa, and the New Guinea highlander. Detailed description of these populations are presented elsewhere (Szathmary et al. 1983; Long et al. 1986; Barrantes et al. 1990; Deka et al. 1995b).

DNA Analysis

PCR analysis of CTG repeats was performed by amplification of 100 ng of genomic DNA in a total volume of 25- μ l reaction mixture containing standard PCR buffer, 200 μ M each dNTP, 1 unit of *Taq* polymerase. The primer sequences are as described in Fu et al. (1992). The forward primer was end-labeled using [γ - 33 P]ATP and T4 polynucleotide kinase. The amplified products were separated on 6% denaturing polyacrylamide gels. Following electrophoresis, the gels were dried and allelic fragments were visualized by autoradiography. Repeat sizes were determined by comparison to an M13 sequence ladder and known controls.

The *Alu* insertion/deletion polymorphisms were detected following the PCR protocol of Mahadevan et al. (1993). One hundred nanograms of DNA was amplified by a hot start PCR incorporating the two forward (405 and 491) and reverse (486) primers in a total volume 25- μ l reaction mixture containing standard PCR buffer, 200 μ M each dNTP, and 1 unit of *Taq* polymerase. The amplified products were separated on a 2% agarose gel containing ethidium bromide.

Statistical Analysis

From the genotypic data on (CTG) repeat sizes and *Alu*(+/-) alleles, we computed the CTG-*Alu* haplotype frequencies by following the algorithm of Long et al. (1995). Disequilibrium coefficients were measured by D/D_{\max} in reference to the *Alu*(+) allele where $D = p_{ab} - p_a p_b$, p_{ab} being the estimated haplotype frequency of a CTG allele under the *Alu*(+) background, p_a = marginal frequency of the specific CTG-repeat allele, and p_b = frequency of *Alu*(+) (Lewontin 1964).

$$D/D_{\max} = \begin{cases} \text{minimum of } p_a(1-p_b) \\ \text{and } (1-p_a)p_b \text{ when } D > 0 \\ \text{minimum of } p_a p_b \text{ and} \\ (1-p_a)(1-p_b) \text{ when } D < 0. \end{cases}$$

Therefore, D/D_{\max} is always between -1 and +1, the limits attaining in cases of exclusive association of the specific CTG repeat allele under *Alu*(-) and *Alu*(+) background, respectively. The significance of the disequilibrium coefficients were judged from the χ^2 statistic $\chi^2 = nD^2 / \Sigma p_a p_b (1-p_a)(1-p_b)$, which follows a χ^2 distribution with 1 df. Values of D/D_{\max} yielding $\chi^2 > 3.84$ are significant at $P < 0.05$ (denoted by *) and $\chi^2 > 6.63$ are significant at $P < 0.01$ (denoted by **) in Tables 2-6.

ACKNOWLEDGMENTS

We thank Drs. P.E. Smouse, J.C. Long, and J.W. Wood for providing the New Guinea samples, Dr. E.J.E. Szathmari for the Dogrib samples, and Dr. G. Flatz for the German samples. This research was supported by grants GM45861, GM41399, and AG09375 from the National Institutes of Health, and by the W.M. Keck Center for Advanced Training in Computational Biology at the University of Pittsburgh and Carnegie Mellon University.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Ashizawa, T. and H.F. Epstein. 1991. Ethnic distribution of myotonic dystrophy gene. *Lancet* **338**: 642-643.
- Barrantes, R., P.E. Smouse, H.W. Mohrenweiser, H. Gershowitz, J. Azofeifa, T.D. Arias, and J.V. Neel. 1990. Microevolution in Lower Central America: Genetic characterization of the Chibcha-speaking groups of Costa Rica and Panama, and a consensus taxonomy based on genetic and linguistic affinity. *Am. J. Hum. Genet.* **46**: 63-84.
- Brook, J.D., M.E. McCurrach, H.G. Harley, A.J. Buckler, D. Church, H. Aburatani, K. Hunter, V.P. Stanton, J-P. Thirion, T. Hudson, R. Sohn, B. Zelman, R.G. Snell, S.A. Rundle, S. Crow, J. Davies, P. Shelbourne, J. Buxton, C. Jones, V. Juvonen, K. Johnson, P.S. Harper, D.J. Shaw, and D.E. Housman. 1992. Molecular basis of myotonic dystrophy: Expansion of trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* **68**: 799-808.
- Buxton, J., P. Shelbourne, J. Davies, C. Jones, T. Van Tongeren, C. Aslandis, P. de Jong, G. Jansen, M. Anvret, B. Riley, R. Williamson, and K. Johnson. 1992. Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. *Nature* **355**: 547-548.
- Carey, N., K. Johnson, P. Nokelainen, L. Peltonen, M.-L. Savontaus, V. Juvonen, M. Anvret, U. Grandell, K. Chotai, E. Robertson, H. Middleton-Price, and S. Malcom. 1994. Meiotic drive at the myotonic dystrophy locus? *Nature Genet.* **6**: 117-118.
- Dada, T.O. 1973. Dystrophia myotonica in Nigerian family. *East Afr. Med. J.* **50**: 213-228.
- Davies, J., H. Yamagata, P. Shelbourne, J. Buxton, T. Ogihara, P. Nokelainen, M. Nakagawa, R. Williamson, K. Johnson, and T. Miki. 1992. Comparison of the myotonic dystrophy associated CTG repeat in European and Japanese populations. *J. Med. Genet.* **29**: 766-769.
- Deka, R., M.D. Shriver, L.M. Yu, R.E. Ferrell, and R. Chakraborty. 1995a. Intra- and inter-population diversity at short tandem repeat loci in diverse populations of the world. *Electrophoresis* **16**: 1659-1664.
- Deka, R., L. Jin, M.D. Shriver, L.M. Yu, S.D. DeCoo, J. Hundrieser, C.H. Bunker, R.E. Ferrell, and R. Chakraborty. 1995b. Population genetics of dinucleotide (dC-dA)_n-(dG-dT)_n polymorphisms in world populations. *Am. J. Hum. Genet.* **56**: 461-474.
- Fu, Y.-H., A. Pizzuti, R.G. Fenwick Jr., J. King, S. Rajnarayan, P.W. Dunne, J. Dubel, G.A. Nasser, T. Ashizawa, P. de Jong, B. Wieringa, R. Korneluk, M.B. Perryman, H.F. Epstein, and C.T. Caskey. 1992. An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* **255**: 1256-1258.

- Gennarelli, M., B. Dallapiccola, M. Baiget, I. Martorell, and G. Novelli. 1994. Meiotic drive at the myotonic dystrophy locus. *J. Med. Genet.* **31**: 980.
- Goldman, A., M. Ramsay, and T. Jenkins. 1994. Absence of myotonic dystrophy in southern African Negroids is associated with a significantly lower number of CTG trinucleotide repeats. *J. Med. Genet.* **31**: 37–40.
- Goldman, A., M. Ramsay, and T. Jenkins. 1995. New founder haplotypes at the myotonic dystrophy locus in Southern Africa. *Am. J. Hum. Genet.* **56**: 1373–1378.
- Harley, H.G., J.D. Brook, S.A. Rundle, S. Crow, W. Reardon, A.J. Buckler, P.S. Harper, D.E. Housman, and D.J. Shaw. 1992. Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. *Nature* **355**: 545–546.
- Harper, P.S. 1989. *Myotonic dystrophy*. Saunders, London, UK.
- Harper, P.S., H.G. Harley, W. Reardon, and D.J. Shaw. 1992. Anticipation in myotonic dystrophy: New light on an old problem. *Am. J. Hum. Genet.* **51**: 10–16.
- Hurst, G.D.D., L.D. Hurst, and J.A. Barrett. 1995. Meiotic drive and myotonic dystrophy. *Nature Genet.* **10**: 132–133.
- Imbert, G., C. Kretz, K. Johnson, J.-L. Mandel. 1993. Origin of the expansion mutation in myotonic dystrophy. *Nature Genet.* **4**: 72–76.
- Jansen, G., M. Mahadevan, C. Amemiya, N. Wormskamp, B. Segers, W. Hendriks, K. O'Hoy, S. Baird, L. Sabourin, G. Lennon, P.L. Jap, D. Iles, M. Coerwinkel, M. Hofker, A.V. Carrano, P.J. de Jong, R.G. Korneluk, and B. Wieringa. 1992. Characterization of the myotonic dystrophy region predicts multiple protein isoform-encoding mRNAs. *Nature Genet.* **1**: 261–266.
- Krahe, R., M. Eckhart, A.O. Ogunniyi, B.O. Osuntokun, M.J. Siciliano, and T. Ashizawa. 1995. De novo myotonic dystrophy mutation in a Nigerian kindred. *Am. J. Hum. Genet.* **56**: 1067–1074.
- Lavedan, C., H. Hofmann-Radvanyi, C. Boileau, C. Bonaïti-Pellié, D. Savoy, P. Shelbourne, C. Duros, J.-P. Rabes, I. Dehaupas, S. Luce, K. Johnson, and C. Junien. 1994. French myotonic dystrophy families show expansion of a CTG repeat in complete linkage disequilibrium with an intragenic 1 kb insertion. *J. Med. Genet.* **31**: 33–36.
- Lewontin, R.C. 1964. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **49**: 49–67.
- Long, J.C., J.M. Naidu, H.W. Mohrenweiser, H. Gershowitz, P.L. Johnson, J.W. Wood, and P.E. Smouse. 1986. Genetic characterization of Gainj- and Kalam-speaking peoples of Papua New Guinea. *Am. J. Phys. Anthropol.* **70**: 75–96.
- Long, J.C., R.C. Williams, and M. Urbanek. 1995. An E-M algorithm and testing strategy for multiple-locus haplotypes. *Am. J. Hum. Genet.* **56**: 799–810.
- Mahadevan, M., C. Tsilfidis, L. Sabourin, G. Shutler, C. Amemiya, G. Jansen, C. Neville, M. Narang, J. Barceló, K. O'Hoy, S. Leblond, J. Earle-Macdonald, P. de Jong, B. Wieringa, and R.G. Korneluk. 1992. Myotonic dystrophy mutation: An unstable CTG repeat in the 3' untranslated region of the gene. *Science* **255**: 1253–1255.
- Mahadevan, M.S., M.A. Foitzik, L.C. Surh, and R.G. Korneluk. 1993. Characterization and polymerase chain reaction (PCR) detection of an *Alu* deletion polymorphism in total linkage disequilibrium with myotonic dystrophy. *Genomics* **15**: 446–448.
- Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583–590.
- Neville, C.E., M.S. Mahadevan, J.M. Barceló, and R.G. Korneluk. 1994. High resolution genetic analysis suggests one ancestral predisposing haplotype for the origin of the myotonic dystrophy mutation. *Hum. Mol. Genet.* **3**: 45–51.
- Rubinsztein, D.C., J. Leggo, W. Amos, D.E. Barton, and M.A. Ferguson-Smith. 1994. Myotonic dystrophy CTG repeats and the associated insertion/deletion polymorphism in human and primate populations. *Hum. Mol. Genet.* **3**: 2031–2035.
- Shriver, M.D., L. Jin, R. Chakraborty, and E. Boerwinkle. 1993. VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. *Genetics* **134**: 983–993.
- Szathmary, E.J.E., R.E. Ferrell, and H. Gershowitz. 1983. Genetic differentiation in Dogrib Indians: Serum protein and erythrocyte enzyme variation. *Am. J. Phys. Anthropol.* **62**: 249–254.
- Watkins, W.S., M. Bamshad, and L.B. Jorde. 1995. Population genetics of trinucleotide repeat polymorphisms. *Hum. Mol. Genet.* **4**: 1485–1491.
- Yamagata, H., T. Miki, T. Ogihara, M. Nakagawa, I. Higuchi, M. Osame, P. Shelbourne, J. Davies, and K. Johnson. 1992. Expansion of unstable DNA region in Japanese myotonic dystrophy patients. *Lancet* **339**: 692.
- Yamagata, H., T. Miki, M. Nakagawa, K. Johnson, R. Deka, and T. Ogihara. 1996. Association of CTG repeats and the 1 kb *Alu* insertion/deletion at the myotonin protein kinase gene in the Japanese population suggests a common Eurasian origin of the myotonic dystrophy mutation. *Hum. Genet.* **97**: 145–147.
- Zerylnick, C., A. Torroni, S.L. Sherman, and S.T. Warren. 1995. Normal variation at the myotonic dystrophy locus in global human populations. *Am. J. Hum. Genet.* **56**: 123–130.

Received November 3, 1995; accepted in revised form January 10, 1996.