A unique origin and multistep process for the generation of expanded DRPLA triplet repeats

Hiroko Yanagisawa1,2, Katsura Fujii1,3, Shigeo Nagafuchi1, Yutaka Nakahori2, Yasuo Nakagome2, Atsushi Akane4, Masataka Nakamura5, Akira Sano6, Osamu Komure7, Ikuko Kondo8, Dong Kyu Jin9, Sven A. Sørensen10, Nicholas T. Potter11, S. Robert Young12, Koicho Nakamura1,13, Nobuyuki Nukina13, Yoshiro Nagao1, Keiko Tadokoro1, Torayuki Okuyama1, Toshiyuki Miyashita1, Tadashi Inoue3, Ichiro Kanazawa13 and Masao Yamada1,*

1National Children’s Medical Research Center, 3-35-31, Taishido, Setagaya, Tokyo 154, Japan, 2Departments of Human Genetics and 13Neurology, University of Tokyo, Tokyo 113, Japan, 3Laboratory of Nucleic Acid Science, Nihon University, Fujisawa 252, Japan, 4Department of Legal Medicine, Kansai Medical University, Moriguchi 570, Japan, 5Human Gene Sciences Center, Tokyo Medical and Dental University, Tokyo 113, Japan, 6Departments of Neuropsychiatry and 8Hygiene, Ehime University School of Medicine, Ehime 791-02, Japan, 7Department of Neurology and Psychiatry, Kobe University School of Medicine, Kobe 650, Japan, 9Department of Pediatrics, Samsung Medical Center, Seoul 135–230, Korea, 10Department of Medical Genetics, IMBG The Panum Institute, Copenhagen N, Denmark, 11Clinical Molecular Genetics Laboratory, The University of Tennessee Medical Center at Knoxville, Knoxville, TN 37920-6999, USA and 12Department of Obstetrics and Gynecology, South Carolina University, Columbia, SC 29203, USA

Received November 6, 1995; Revised and Accepted December 14, 1995

Dentatorubral and pallidolysian atrophy (DRPLA) is an autosomal dominant neurodegenerative disorder associated with the expansion of a CAG repeat at chromosome band 12p13. Epidemiological studies have demonstrated an increased prevalence of DRPLA in Japan, although several DRPLA kindreds of non-Japanese ancestry have been identified. To define the molecular basis for this geographic variation in prevalence, we have analyzed haplotypes around the repeat in several different ethnic groups. Two intragenic biallelic polymorphisms distinguished three haplotypes, each of which formed a predominant haplotype found in the three major racial populations. All the expanded repeats of Japanese and Caucasian patients studied were associated with a particular haplotype, which otherwise was associated with longer repeats commonly found in Asians. Our results support a multi-step model for repeat expansion, and suggest that expanded DRPLA repeats may have evolved from an ancient chromosomal haplotype of Asian origin. We also propose that a combination of a highly polymorphic microsatellite marker with relatively stable biallelic markers in a range of PCR amplification is a powerful tool for studies on human genome diversity, which may reveal the ancient human migration and the formation of ethnic groups.

INTRODUCTION

Dentatorubral and pallidolysian atrophy (DRPLA) is one of the growing number of genetic diseases which are caused by expansion of triplet repeats (1–4). Currently, this group includes spinal and bulbar muscular atrophy (AR gene for SBMA) (5), the fragile X syndromes (FRAXA, FRAXE and FRAXF) (6–10), myotonic dystrophy (DM) (11–14), Huntington’s disease (HD) (15), spinocerebellar ataxia type 1 (SCA1) (16) and Machado-Joseph disease (MJD1) (17). Repeats for DRPLA, HD, SCA1, MJD1 and SBMA consist of a reiterated CAG trinucleotide and share some of the following molecular features: the repeats are situated within a coding region and are translated as a polyglutamine track in the respective protein product (18–20). The repeats on normal chromosomes are usually within a range up to 35 repeats, and are highly polymorphic in the normal population. In contrast, the repeats on disease chromosomes are in a range of 40–100 repeats, and display somatic mosaicism and inter-generational instability particularly in paternal transmission. Repeat
instability is thus common, but distinctive neuropathologic changes are found in each CAG repeat disorder.

Although a sporadic case of DRPLA was originally reported in Europe (21), it has been predominantly described in patients and kindreds of Japanese ancestry (1,2). Because of its clinical heterogeneity ranging from a pseudo-Huntington type in adult-onset cases to a myoclonus epilepsy type in juvenile-onset cases, DRPLA had been very difficult to diagnose without postmortem pathological examinations. This phenomenon, now understood as genetic anticipation, has been well accounted for by a further increase of expanded repeats and inverse correlation of age of onset with the repeat number. Genetic anticipation is especially evident in Japanese DRPLA families, where the size of repeat increases by 8.0 with a decrease by 34.4 years in age of onset, as an average, in paternal transmission (22). Recent retrospective and prospective molecular studies estimate an incidence of two to four per million, a rate that is almost similar to that of HD in Japan. Outside Japan, however, DRPLA remains rare with the current molecular literature reflecting the identification of only seven kindreds (23–26), including one African-American family with Haw River syndrome (HRS) (27). Despite the demonstration of expansion of the same repeat in all patients studied to date, there are several clinical and neuropathological differences between DRPLA in Japan and Western countries which may reflect genetic or environmental effects on the expression of the phenotype (25,27). This includes the absence of myoclonus epilepsy and the presence of extensive subcortical white demyelination in patients with HRS and in some of the Western DRPLA families, and the more dramatic effects of genetic anticipation in Japanese DRPLA families.

To study the difference in ethnic prevalence of DRPLA, we have previously examined the distribution of DRPLA repeat number on Japanese, Caucasian and African-American chromosomess (28). That study demonstrated that the frequency of relatively longer repeats in the normal range for each ethnic group correlated with the prevalence of DRPLA in the population. Here, we extend these analyses using new polymorphic markers to define haplotypes associated with normal and expanded DRPLA repeats. Our study indicated a multi-step process for generation of expanded DRPLA repeats.

RESULTS

Polymorphic markers around the DRPLA repeats

Our initial mapping indicated that the DRPLA gene was located near F8VWF and CD4 on the short arm of chromosome 12. Recently, the gene was mapped by FISH to 12p13.31 (29), and also by YAC analyses to a 6 cM region (30) between D12S93 and D12S77 (Takeda and Yamada, in preparation). However, no particular haplotypes defined by these nearby markers segregated with expanded repeats in Japanese DRPLA families (Table 1). Therefore, we looked for new polymorphic markers utilizing sequence information derived from the full cDNA sequence and partial genomic sequences of the DRPLA gene (31) (see Materials and Methods). Two intragenic biallelic polymorphisms turned out to be informative in the haplotype study. One of the polymorphic sites (the A system) resulted from a single nucleotide substitution in intron 1, and could be detected by PCR–SSCP. The other (the B system) was situated in intron 3, and could be detected by either PCR–SSCP or Eco811I digestion of the PCR products (Figs 1 and 2).

<table>
<thead>
<tr>
<th>Haplotypes associated with expanded DRPLA repeats when defined with polymorphic markers outside the DRPLA gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Families</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>D</td>
</tr>
<tr>
<td>E</td>
</tr>
<tr>
<td>F</td>
</tr>
<tr>
<td>G</td>
</tr>
<tr>
<td>H</td>
</tr>
<tr>
<td>I</td>
</tr>
<tr>
<td>J</td>
</tr>
</tbody>
</table>

The allele number is assigned in a descending order of fragment size detected in this analysis.

The DRPLA gene was situated between F8VWF and CD4 (unpublished result). Recombinations between F8VWF and the CAG repeat were observed in some descendants. Only the ancestral haplotype is indicated.

The associated allele was not determined at several sites because of a small pedigree of two generations.

The allele number and frequency in parenthesis are given for alleles frequently observed in spouses of the families in this analysis.

**Haplotype analysis in Japanese DRPLA families**

Eighty-eight independent chromosomes from 93 members of 17 DRPLA families were studied. We first analyzed the A, B and CAG repeat polymorphisms in separate PCR amplification, and then determined the haplotype of each chromosome based on a segregation pattern. The frequency of each haplotype and its association with CAG repeat numbers are illustrated in Figure 3.

Although four haplotypes were possible with two biallelic polymorphic systems, only three were observed. As previously reported (3) and also described below, numbers of the CAG repeat were distributed from 7 to 23 in the normal Japanese population with three peaks at 10, 15 and 17 repeats. Familial studies indicated that repeats at about 10, 15 and 17 were essentially associated with the A2–B2, A1–B2 and A1–B1 haplotypes, respectively. All the expanded alleles (>49 repeats) were segregated with the A1–B1 haplotype.

**Haplotype analysis of normal chromosomes**

We then analyzed the three polymorphisms in DNA samples isolated from normal individuals of various ethnic groups, i.e. Japanese, Korean, Chinese, Caucasian and African-American. When samples showed homozygosity at two of the three polymorphic sites, haplotypes were easily assigned. Samples heterozygous at the B polymorphic site were subjected to allele-specific-oligonucleotide primed PCR (32) to determine the number of repeats on each chromosome. Some of the double-heterozygous samples were also subjected to amplification of the entire region encompassing the two or three polymorphic sites by long PCR (33), and the phased haplotype was determined using
Figure 1. Schematic illustration of the polymorphisms used for haplotype analyses. The top line shows approximate positions of 10 exons of the DRPLA gene (ref. 30 and unpublished result). The second line shows the position of the three polymorphic sites used in this study (system A, system B, and CAG repeat). PCR primer pairs and the size of the products are shown below the line which were used to detect each polymorphism, to determine the association of alleles by allele-specific-oligonucleotide-primed PCR and to clone DNA fragments covering more than two polymorphic sites. The polymorphic A system resulted from a single nucleotide substitution at the 1010th position of the 2070 bp BglII–BglII fragment. The B polymorphic system was resulted from a single nucleotide substitution at the 1865th position of the fragment, and could be distinguished by digestion with Eco81I. The revised cDNA sequence and the genomic sequence of the 2070 bp BglII–BglII fragment have been deposited in the DDBJ, GSDB, EMBL and NCBI nucleotide databases with the accession numbers of D31840 and D63808, respectively.

Figure 2. Analysis of the PCR-amplified products containing the A and B polymorphic sites. (A) shows a PCR-SSCP pattern for the A polymorphism detected by a mini-gel system (PhastSystem, Pharmacia). Lanes 3–5 were A1/A1 homozygotes, lanes 2 and 6 were A1/A2 heterozygotes, and lanes 1 and 7 were A2/A2 homozygotes. (B) shows a PCR-SSCP pattern for the B polymorphism detected by the same system. Lanes 1, 2 and 4 were B1/B1 homozygotes, lanes 3 and 6 were B1/B2 heterozygotes and lane 5 was a B2/B2 homozygote. (C) shows a result of gel electrophoresis of the same PCR products as (B) after Eco81I digestion. Size markers (M) are Sau3A-digested pUC18.

cloned fragments. Even after extensive analyses, no A2–B1 haplotypes were detected in any samples so far analyzed, suggesting that the B1 allele was exclusively linked to the A1 allele.

The distribution of CAG repeat number together with the frequency of the B1 allele (thus the A1–B1 haplotype) in these ethnic groups is shown in Figure 4. On Japanese chromosomes, CAG repeats formed a trimodal distribution with peaks at 10, 15 and 17 repeats. All the repeats longer than 17, including repeats in an intermediate range, were exclusively associated with the A1–B1 haplotype. In Korean and Chinese chromosomes, the
Figure 4. Distribution of CAG repeat number with an associated haplotype detected in individuals of six ethnic groups. The number of chromosome studied and the frequency of the three haplotypes are indicated. filled bars; A1–B1, hatched bars; A1–B2, blank bars; A2–B2.

distribution of repeat number and associated haplotypes were similar to that observed in Japanese, but (CAG)_{15} was the most frequently observed normal allele. This was accounted for by the lower frequency of the A2 allele and the increase of the A1–B2 haplotype in these groups. In general, about half of Asian chromosomes had the A1–B1 haplotype which was usually associated with longer repeats in the normal range. Although the upper tail in the repeat distribution including intermediate repeats were more prominent in Japan, one Chinese chromosome with 33 repeats was identified, suggesting a possible occurrence of DRPLA in other Asian countries.

Repeat polymorphism in Caucasians showed smaller heterozygosity in a narrow range, as we had previously shown with another set of samples. Double homozygous and representative cloned samples showed that repeats around 10 and 15 were associated with the A2–B2 and A1–B2 haplotypes, respectively, which was similar to that observed in Asian populations. Only 8% of the chromosomes had the A1–B1 haplotype with repeats up to 19, while 66% had the A1–B2 haplotype with repeats around 15. Unlike Asians and Caucasians, the predominant haplotype (77%) on African chromosomes was A2–B2, and widely distributed in a range of eight to 17 repeats. Only 4% of chromosomes in this group had the A1–B1 haplotype.

To investigate anthropologic questions regarding the introduction of the A1–B1 haplotype into the Japanese population, we conducted a similar analysis on a Japanese subgroup, carriers of the human T-lymphotrophic virus (HTLV-1), the causative agent of adult T-cell leukemia (ATL). ATL is endemic in a defined area including the Kyushu island located southwest of Japan as well as in central and western Africa and also in the Caribbean. As phylogenetic studies on the isolated virus genome suggested HTLV-1 was introduced into Japan in an ancient period with movements of Mongoloids, thus HTLV-1 carriers in Japan are thought to represent the aboriginal inhabitants of the Japanese island (Jomon-jin) (34–36). The distribution of DRPLA alleles and associated haplotypes in the HTLV-1 carriers are similar to those found on other Japanese chromosomes.

Haplotype analyses of DRPLA patients in Western countries

We analyzed DNA samples from two DRPLA families identified in Western countries. Expanded repeats found in a large pedigree of a Danish family (24) and another North American Caucasian family (25) were associated with the A1–B1 haplotype (data not shown). As far as the family members disclosed, the family history indicated no evidence of Asian ancestry.

Haplotype analyses of rodent's and primate's genome

The regions covering the polymorphic A, B and CAG repeats were evolutionarily conserved. The primer sets used to detect the polymorphic sites in human genome also amplified comparative DNA fragments from genomes of rodents and primates, except for the A polymorphic site of rodents. Repeats in these species were generally smaller, for examples, CAGCAA(CAG)_{5} CAA(CAG)_{7} and CAACAGCAA(CAG)_{9} in representative clones of orang-utan (Pongo pygmaeus) and chimpanzee (Pan troglodytes), respectively. All the rodents and primates analyzed had an EcoB11 site (thus the B2 allele) and all the primates analyzed a G nucleotide (thus the A2 allele) at the corresponding positions.

DISCUSSION

Triplet repeat expansion is a novel type of mutational mechanism for human genetic diseases and has now provided a molecular explanation for the concept of genetic anticipation. Yet, two fundamental questions remain to be answered. First, what is the mechanism responsible for instability of expanded repeats? Second, how do expanded repeats lead to neuron death, especially in CAG repeat disorders, in a specific anatomical area? Haplotype analyses of normal and expanded repeats can provide some data to address this first question. However, it will remain difficult to distinguish between cis-acting elements and a founder effect when these unstable alleles share a common haplotype.

Haplotypes on DM, HD, FRAXA and MJD1 chromosomes have been studied. Various polymorphic analyses had suggested
a common haplotype around the disease gene during mapping studies for the HD and DM loci. With a number of polymorphic sites in the DM gene finally identified, all the DM chromosomes found in various ethnic groups had been shown to once share a common haplotype, which indicated a unique origin of expanded DM repeats (37). A recent study, however, has indicated another common haplotype, which indicated a unique origin of expanded disease alleles in various ethnic groups had been shown to once share a unique haplotype, A1–B1. This haplotype is prevalent in Asian populations and usually associated with longer repeats in the normal range, but the upper trail in the distribution of repeat number is most prominent in the Japanese population. This provides genotype data in support of the increased frequency of DM repeat in Japan. Although two biallelic polymorphisms have a limited power to distinguish haplotypes, nevertheless, they were able to distinguish the predominant haplotype associated with three major racial populations. The particular haplotype was also found on chromosomes of the Caucasian DRPLA patients, while it rarely occurred on normal Caucasian chromosomes. These results suggest that Japanese and Caucasian DRPLA disease alleles may have been derived from a common founder chromosome.

The A2–B2 haplotype seems to be an ancestral form in mammalian evolution on the basis of results in rodents and primates. A mutation to create the A1 allele occurred on a chromosome having about 15 repeats, and then another mutation to create the B1 allele occurred on a chromosome with the A1 background and about 17 repeats (Fig. 5). This is consistent with the finding that only three haplotypes among the four possible combinations have been detected in our study. As the A1–B1 haplotype is dominant in the Asian population, the mutation to create the B1 allele seems to have occurred in this ethnic group. Japanese HTLV-1 carriers, who represent aboriginal inhabitants of the Japanese island (34–36), also showed a similar haplotype distribution. Taken together, these results suggest that descendants having the A1–B1 haplotype with longer repeats within a normal range may have come to Japan in ancient periods with the movement of Mongoloids, and repeat number may have further increased occasionally in a genetically isolated environment. This is partially supported by the data demonstrating that DRPLA chromosomes displayed a variety of haplotypes when defined with markers separated by a few centimorgans from the repeat. Precursor forms of DRPLA chromosomes may also have been introduced into the Caucasian population in ancient periods probably by admixtures with Asian ancestors, although alternative explanations may be possible. To resolve this question, haplotype analyses in a wide range of ethnic groups would be required. Of particular interest would be studies of Native American populations, who are thought to be descendants of Mongoloids, and the Central and South Asian ethnic groups. Our study found relatively long repeats unassociated with the A1–B1 haplotype in Africans. Other studies showed a few cases of relatively longer repeats (21 and 29 repeats) in Africans (28,42).
The PCR products showed different mobilities in SSCP analyses. The variation was resulted from a single nucleotide substitution at the 1865th position of the BglII-BglII sequence. As this site was covered by a recognition sequence of the EcoR11 restriction enzyme, the B allele could also be detected by digestion of the enzyme. We used a standard PCR condition previously described (49) unless otherwise specified. Primers, annealing temperature and concentration of MgCl2 were adjusted as follows: A polymorphism; DALUF, 5′-CTTGTTGAGTGGGCTGAGTCA; DALUR, 5′-AGCTATC- TAAAAGGGAGGCGAATTT; at 60°C and 1.5 mM. B polymorphism; DGTF, 5′-GTGATGCGAAAGCTGTGAGAAGT; DGTR, 5′-CCGTCTTGTTGACCTTTGGGG; at 60°C and 1.5 mM.

Allele-specific-oligonucleotide primed PCR (32)

To determine the phase of the B and CAG repeat alleles in B1/B2 heterozygous samples, two PCR conditions were set up where either a primer combination of EcoR11T with DR2124, or another primer combination of EcoR11C with DR2124 was included. The PCR conditions amplified about 2500 bp covering the CAG repeat. Then, numbers of DRPLA repeats were measured for each reaction product with a CTG-B37 primer set as described above. PCR conditions used in the first amplification were essentially the same as our standard condition with annealing temperature at 60°C and MgCl2 concentration of 1.5 mM, except for additional inclusion of 1.25 U of Taq polymerase (Takara), and then cloned on an Sma1 sequence. As this site was covered by a recognition sequence of the EcoR11 restriction enzyme, the B allele could also be detected by digestion of the enzyme. We used a standard PCR condition previously described (49) unless otherwise specified. Primers, annealing temperature and concentration of MgCl2 were adjusted as follows: A polymorphism; DALUF, 5′-CTTGTTGAGTGGGCTGAGTCA; DALUR, 5′-AGCTATC- TAAAAGGGAGGCGAATTT; at 60°C and 1.5 mM. B polymorphism; DGTF, 5′-GTGATGCGAAAGCTGTGAGAAGT; DGTR, 5′-CCGTCTTGTTGACCTTTGGGG; at 60°C and 1.5 mM.

Long PCR

To determine phase for more than two alleles, DNA fragments were amplified by long PCR (33), treated with T4 DNA polymerase (Takara), and then cloned on an Smal site of pUC18. Each allele was then determined on isolated clones, usually on independent six clones. PCR conditions used in long PCR were described above with the following oligonucleotide primers in addition to DGTR, DALUF and CTG-B37R: DUALUF, 5′-AGGTATGCTTGGGATTTG; annealing temperature for PCR with DUALUF and DGTR was 60°C, while that with DALUF and CTG-B37R was 65°C.

Other techniques

DNA preparation from leukocytes and plasmid isolation were carried out following standard methods as described (3,30,50). Nucleotide sequences were determined using a Sequenase version 2 kit (USB) with universal primers in pUC18 or internal primers. Images with radioactivity were detected and quantified using a Bioimage Analyzer BAS-2000 (Fuji).


