Analysis of spinocerebellar ataxia type 2 gene and haplotype analysis: $(CCG)_{1-2}$ polymorphism and contribution to founder effect

Kazuyuki Mizushima, Mitsunori Watanabe, Ikuko Kondo, Koichi Okamoto, Masami Shizuka, Koji Abe, Masashi Aoki, Mikio Shoji

Abstract

Spinocerebellar ataxia type 2 is a familial spinocerebellar ataxia with autosomal dominant inheritance. The gene responsible was recently cloned and this disorder was found to be the result of a CAG expansion in its open reading frame. We analysed 13 SCA2 patients in seven unrelated families in Gunma Prefecture, Japan. In four of the seven families, we detected CCG or CCGCCG interruptions in only the expanded alleles. Cosegregation of these polymorphisms with SCA2 patients was established within each family. Together with the results of haplotype analyses, we considered that at least two founders were present in our area and that these $(CCG)_{1-2}$ polymorphisms may make analysis of founder effects easier. By sequencing analysis we found that although the number of the long CAG repeat varied in each subclone of expanded alleles, these polymorphisms did not change their configuration. This finding suggests that CCG or CCGCCG sequences are stable when surrounded by the long CAG repeat and a single CAG. Moreover, the presence of these polymorphisms may lead to miscounting the repeat size by conventional estimation using a size marker such as an M13 sequencing ladder. Therefore we should consider these polymorphisms and accurately determine the repeat size by sequencing.

(*J Med Genet* 1999;**36**:112–114)

Keywords: spinocerebellar ataxia type 2; CCG repeat polymorphism; founder effect

Department of Neurology, Gunma University School of Medicine, 3-39-22 Showa-machi, Maebashi 371-8511, Japan
K Mizushima
M Watanabe
K Okamoto
M Shizuka
M Shoji

Department of Hygiene, Ehime University School of Medicine, Shizukawa, Shigenobu-cho, Onsen-gun 791-0295, Japan
I Kondo

Department of Neurology, Tohoku University School of Medicine, 1-1 Seiryo-machi, Aoba-ku, Sendai 980-8574, Japan
K Abe
M Aoki

Correspondence to:
Dr Watanabe.

Received 26 November 1997
Revised version accepted for publication 1 July 1998

The autosomal dominant spinocerebellar ataxias (ADCA) consist of genetically heterogeneous subtypes.[1] The gene responsible for spinocerebellar ataxia type 2 (SCA2), one type of ADCA, was recently cloned, and SCA2 was shown to occur as a result of the expansion of an unstable CAG repeat in its open reading frame,[2–4] similar to SCA1,[5] SCA3 allelic to Machado-Joseph disease (MJD),[6] SCA6,[7] SCA7,[8] hereditary dentatorubropallidoluysian atrophy (DRPLA),[9 10] Huntington's disease (HD),[11] and X linked spinobulbar muscular atrophy (SBMA).[12] Previous reports have shown that CAA interruption ranging from one to three is only present in the normal alleles of the SCA2 gene, and this interruption possibly stabilises replication of the normal alleles.[2–4] However, there have been no reports concerning the interruptions observed in the expanded CAG repeat of the SCA2 gene. In this study, we document two types of previously unreported polymorphisms present only in the expanded alleles and their contribution to easier analysis of the founder effect in association with data from haplotyping.

Materials and methods

Thirteen SCA2 patients (five males and eight females) from seven unrelated Japanese families (families A-G) and 60 unrelated normal controls who live in Gunma Prefecture (located about 100 km north of Tokyo with a population of approximately two million) were genetically examined. All patients had a family history of autosomal dominant inheritance. Blood samples were obtained from these cases with informed consent. Genomic DNA was extracted from blood by a standard phenol/chloroform method. Using a primer set of SCA2-A whose 5' end was Cy5 labelled and SCA2-B from Pulst's group, the region containing the CAG repeat in the SCA2 gene was amplified by polymerase chain reaction (PCR) according to previously reported methods.[3] After the PCR products were analysed by 4% agarose gel electrophoresis with ethidium bromide staining, an aliquot of the PCR product was electrophoresed on 6% denaturing polyacrylamide gel with an automated DNA sequencer (ALF Express, Pharmacia LKB, Uppsala, Sweden). The data were processed with fragment analysis software (Fragment Manager, Pharmacia) according to our previous methods.[13–15] After confirming the presence of both normal and expanded alleles on agarose and sequencing gels, both alleles were excised from the agarose gel and purified by spin columns (GenElute agarose spin columns, Supelco, Bellefonte, USA). Thereafter these purified fragments were subcloned into a plasmid vector (pBluescript II KS (+), Stratagene, La Jolla, CA, USA) and the sequence was determined with the above sequencer. All fragments were analysed by plasmid sequencing, while downstream of the reverse primer (SCA2-B) was not examined in this study. Similarly, analyses of six microsatellite markers (D12S84, D12S105, D12S1672, D12S1333, D12S354, and D12S79) flanking the SCA2 gene (cen-D12S84/D12S105-1 cM-D12S1672 (at the first intron in the SCA2 gene)-20 kb-the CAG repeat in the SCA2 gene-200 kb-D12S1333-4 cM-D12S354-2

J Med Genet: first published as 10.1136/jmg.36.2.112 on 1 February 1999. Downloaded from http://jmg.bmj.com/ on June 5, 2023 by guest. Protected by copyright.

Table 1 Results of sequencing and haplotype analysis in 13 patients from seven SCA2 families

Family	*Case*	*Relation to other cases*	*Repeat size*	*Config of poly*	*D12S84*	*D12S105*	*D12S1672*	*D12S1333*	*D12S354*	*D12S79*
A	1 M	Father of A-2	37, 37	Normal	10	2	4	7	3	8
	2 F	Proband	44	Normal	10	2	4	7	3	8
B	1 F	Proband	51, 53	Normal	8 or 10	1 or 2	2 or 3	2 or 6	6	2 or 3
C	1 F	Proband	43, 44, 46	Normal	9 or 10	1 or 2	1 or 2	5 or 6	4 or 6	3 or 5
D	1 F	Older sister of D-2	37	CCG	8	2	2	2	1	5
	2 F	Proband	41, 42	CCG	8	2	2	2	1	5
E	1 M	Father of E-2	37, 37, 37	CCG	3	2	2	2	1	3 or 7
	2 M	Proband	44	CCG	3	2	2	2	1	3 or 7
F	1 F	Older sister of F-2	41	CCGCCG	7	2	2	2	6	5
	2 M	Uncle of F-3	42, 42, 43	CCGCCG	7	2	2	2	6	5
	3 F	Proband	51, 51	CCGCCG	7	2	2	2	6	5
G	1 F	Older sister of G-2	37, 38, 40	CCGCCG	7	2	2	2 or 6	3 or 6	5
	2 M	Proband	36	CCGCCG	7	2	2	2 or 6	3 or 6	5

Repeat size means the number of long CAG repeats in each subclone of the expanded alleles.
Config of poly: configuration of polymorphisms. F: female. M: male.

cM-D12S79-tel) were performed and the haplotype of each marker was determined.[16 17] Each allele was numbered depending on its product size. The difference in allele distribution between disease and normal chromosomes was evaluated by the χ^2 and Fisher's exact tests. The Yates correction was used when appropriate. Only phase known genotypes were included in the statistical analysis.

Results

Plasmid sequencing of both the expanded and normal alleles was performed in the above 13 SCA2 patients and 15 normal controls. As in previous reports,[2–4 17] of the 30 normal alleles examined, 25 displayed $(CAG)_8CAA(CAG)_4CAA(CAG)_8$, three $(CAG)_{13}CAA(CAG)_8$, and two $(CAG)_8CAA(CAG)_4CAA(CAG)_7$. However, CCG polymorphisms or other insertions, deletions, or mutations were not observed. As shown in table 1, the sequence of the expanded allele in families A, B, and C had no interruptions after the CAG repeats, whereas that of families D and E had a CCG interruption surrounded by the long CAG repeat and a single CAG, and that of families F and G had a CCGCCG interruption. The expanded alleles of the PCR products on polyacrylamide gel electrophoresis showed multiple peaks. The multiplicity of peaks for the expanded allele is not uniform, suggesting heterogeneity of genomic template in SCA2 expanded alleles. All the PCR products from subcloned expanded DNA, normal alleles, or microsatellite markers tested showed an absence of peaks after the strongest, indicating that the multiple peaks of the expanded alleles for these templates are not the result of *Taq* slippage. Therefore, several subclones were sequenced in each expanded allele. Consequently, variability was only seen in the long CAG repeat and not in the CCGCAG or the CCGCCGCAG (table 1). Moreover, cosegregation of SCA2 patients with an expanded SCA2 gene with or without these polymorphisms was established within each family (table 1).

The above six microsatellite markers were typed in all patients in these seven families and 60 normal controls (tables 1 and 2). As table 1 shows, family A has a different haplotype from families D-G. Families D-G had allele 2 of D12S1672 and families D-F (maybe family G also) possessed allele 2 of D12S1333, both of which were significantly more frequent in patients than in controls ($p<0.05$). Therefore, at least one founder with allele 2 of D12S1672 and allele 2 of D12S1333 is thought to be present in families D-G. In addition, since the families with the CCG polymorphism (families D and E) had the haplotype 2-2-2-1 (D12S105-D12S1672-D12S1333-D12S354), and those with the $(CCG)_2$ polymorphism (families F and G) had the haplotype 7-2-2-5 (D12S84-D12S105-D12S1672-D12S79) and maybe 2-6 (D12S1333-D12S354), it is possible that the founder of family D is the same as that of family E, and family F shares the same founder with family G. Furthermore, families D-F (maybe family G also) may have the same founder because they bear the haplotype 2-2-2 (D12S105-D12S1672-D12S1333). A third founder may be present in families B and C. Taken altogether, at least two founders (one founder for family A and the other for families D-G) are thought to be present in our area.

Table 2 Distributions of chromosome 12 markers in SCA2 patients and normal controls

Locus	*Allele*	$n_{patients}/N$	$n_{controls}/N$	χ^2	*p*
D12S84	3	1/5	6/120		
	7	2/5	18/120		
	8	1/5	30/120		
	10	1/5	10/120		
D12S105	2	5/5	61/120		
D12S1672	2	4/5	18/120	9.86	<0.05
	4	1/5	27/120		
D12S1333	2	3/4	8/120	14.71	<0.001
	7	1/4	31/120		
D12S354	1	2/5	43/120		
	3	1/5	17/120		
	6	2/5	39/120		
D12S79	5	3/4	15/120	7.67	<0.05
	8	1/4	4/120		

N = number of families in which the allele segregating with the mutation could be determined or number of control chromosomes used.

Discussion

In this study, we sequenced the region containing the CAG repeat in the SCA2 gene from seven unrelated SCA2 families in Gunma Prefecture, Japan. In two of them (families D and E), we detected the CCG polymorphism after the long CAG repeat, and in another two (families F and G) the

J Med Genet: first published as 10.1136/jmg.36.2.112 on 1 February 1999. Downloaded from http://jmg.bmj.com/ on June 5, 2023 by guest. Protected by copyright.

CCGCCG configuration was found only in the expanded allele. So far a similar interruption has not been reported in the genes responsible for other CAG related disorders. Moreover, cosegregation of SCA2 patients with the expanded SCA2 gene with or without these polymorphisms was established within each family (table 1). Together with the results of haplotype analyses (tables 1 and 2) it was determined that at least two founders (family A and families D-G) are present in our area. If the founder of families D-G is the same, the CCG repeat has been ancestrally lost or gained through transmissions. From the results of haplotype analyses, families D-G obviously have a different founder from that of family A. Therefore, without analysis of flanking microsatellite markers, we could study the presence of a distinct founder in SCA2 families with the $(CCG)_{1\text{-}2}$ polymorphisms. Thus the $(CCG)_{1\text{-}2}$ polymorphisms may make analysis of founder effects easier. The fact that two founders are present in a geographically confined region suggests that the CAG expansion is prone to occur in the SCA2 gene in normal populations, which is compatible with a recent report.[17] This may be associated with the relatively smaller size of the CAG expansion in the SCA2 gene.

Although the single or double CCG repeat was surrounded by the long CAG repeat and a single CAG, the long CAG repeat which skipped the CCG or CCGCCG sequence was never observed in any of the expanded alleles examined (table 1). Moreover, although the number of the long CAG repeat varied in each subclone of the expanded alleles, these polymorphisms did not change their configuration. The present results clearly show that the $(CCG)_{1\text{-}2}$ sequence is stable when surrounded by the long CAG repeat and a single CAG. The data presented here may provide clues for elucidating the mechanisms of the expansion of trinucleotide repeats.

Because the multiple band pattern of the expanded alleles on polyacrylamide gel electrophoresis suggested that cell mosaicism occurs in the expanded SCA2 gene as well, several subclones were sequenced in each expanded allele. As described above, in contrast with the size variation of the long CAG repeat in each subclone, the CCG and CCGCCG polymorphisms and the single CAG after these polymorphisms never changed their configurations, suggesting that cell mosaicism in the expanded alleles is the result of the variation in the size of the long CAG repeat. In addition, in HD it has been noted that the CCG region immediately 3' to the unstable CAG repeats in the IT15 cDNA is polymorphic, but only the CAG repeat is associated with its expansions and the disease.[18–20] These findings suggest that the size of the long CAG repeat, not the size of the following $(CCG)_{1\text{-}2}$ CAG, is probably responsible for the occurrence of the disease in our patients (families D-G). In CAG related disorders, the size of the expanded allele is conventionally determined on polyacrylamide gel electrophoresis using a size marker such as an M13 sequencing ladder. Therefore, in analyses of the expanded alleles containing these polymorphisms, two or three repeats may be overestimated. Thus, we should examine in detail whether such polymorphisms are contained in the expanded alleles of the genes for other triplet repeat diseases as well as other group SCA2 families.

The first two authors contributed equally to this study. We are indebted to the families studied for their cooperation and support. We also thank Dr Y Igeta for collecting blood samples and Dr Y Ikeda for his support. This work was partly supported by The Nakabayashi Trust For ALS Research, The Uehara Memorial Foundation, and a Grant in Aid for Scientific Research (B) 09470151.

1 Rosenberg RN. Autosomal dominant cerebellar phenotypes: the genotype has settled the issue. *Neurology* 1995;**45**:1-5.
2 Pulst SM, Nechiporuk A, Nechiporuk T, *et al*. Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nat Genet* 1996;**14**:269-76.
3 Imbert G, Saudou F, Yvert G, *et al*. Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. *Nat Genet* 1996;**14**: 285-91.
4 Sanpei K, Takano H, Igarashi S, *et al*. Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique (DIRECT). *Nat Genet* 1996;**14**:277-84.
5 Orr HT, Chung M, Banfi S, *et al*. Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nat Genet* 1993;**4**:221-6.
6 Kawaguchi Y, Okamoto T, Taniwaki M, *et al*. CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nat Genet* 1994;**8**:221-8.
7 Zhuchenko O, Bailey J, Bonnen P, *et al*. Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the α_{1A}-voltage-dependent calcium channel. *Nat Genet* 1997;**15**:62-9.
8 David G, Abbas N, Stevanin G, *et al*. Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. *Nat Genet* 1997;**17**:65-70.
9 Koide R, Ikeuchi T, Onodera O, *et al*. Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). *Nat Genet* 1994;**6**:9-13.
10 Nagafuchi S, Yanagisawa H, Sato K, *et al*. Dentatorubral and pallidoluysian atrophy: expansion of an unstable CAG trinucleotide on chromosome 12p. *Nat Genet* 1994;**6**:14-18.
11 The Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 1993;**72**:971-83.
12 La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* 1991; **352**:77-9.
13 Aoki M, Abe K, Kameya T, Watanabe M, Itoyama Y. Maternal anticipation of DRPLA. *Hum Mol Genet* 1994;**3**: 1197-8.
14 Watanabe M, Abe K, Aoki M, *et al*. Analysis of CAG trinucleotide expansion associated with Machado-Joseph disease. *J Neurol Sci* 1996;**136**:101-7.
15 Watanabe M, Abe K, Aoki M, *et al*. Mitotic and meiotic stability of the CAG repeat in the X-linked spinal and bulbar muscular atrophy gene. *Clin Genet* 1996;**50**:133-7.
16 Allotey R, Twells R, Cemal C, *et al*. The spinocerebellar ataxia 2 locus is located within a 3-cM interval on chromosome 12q23-24.1. *Am J Hum Genet* 1995;**57**:185-9.
17 Geschwind DH, Perlman S, Figueroa CP, Treiman LJ, Pulst SM. The prevalence and wide clinical spectrum of the spinocerebellar ataxia type 2 trinucleotide repeat in patients with autosomal dominant cerebellar ataxia. *Am J Hum Genet* 1997;**60**:842-50.
18 Maddox J. Triplet repeat genes raise questions. *Nature* 1994;**368**:685.
19 Andrew SE, Goldberg YP, Theilmann J, Zeisler J, Hayden MR. A CCG repeat polymorphism adjacent to the CAG repeat in the Huntington disease gene: implications for diagnostic accuracy and predictive testing. *Hum Mol Genet* 1994;**3**:65-7.
20 Barron LH, Rae A, Holloway S, Brock DJH, Warner JP. A single allele from the polymorphic CCG rich sequence immediately 3' to the unstable CAG trinucleotide in the IT15 cDNA shows almost complete disequilibrium with Huntington's disease chromosomes in the Scottish population. *Hum Mol Genet* 1994;**3**:173-5.

J Med Genet: first published as 10.1136/jmg.36.2.112 on 1 February 1999. Downloaded from http://jmg.bmj.com/ on June 5, 2023 by guest. Protected by copyright.