Special article
Arq Neuropsiquiatr 2009;67(4):1124-1132

SPINOCEREBELLAR ATAXIAS

Microsatellite and allele frequency in unaffected and affected individuals

Aline Andrade Freund, Rosana Hermínia Scola, Hélio A.G. Teive, Raquel Cristina Arndt, Magda Clara Vieira da Costa-Ribeiro, Lupe Furtado Alle, Lineu Cesar Werneck

Abstract – The diagnosis and incidence of spinocerebelar ataxias (SCA) is sometimes difficult to analyze due the overlap of phenotypes subtypes and are disorders of mutations caused by CAG trinucleotide repeat expansion. To investigate the incidence of the SCA in Southern Brazil, we analyzed the trinucleotide repeats (CAG)n at the SCA1, SCA2, SCA3, SCA6 and SCA7 *loci* to identify allele size ranges and frequencies. We examined blood sample from 154 asymptomatic blood donors and 115 individuals with progressive ataxias. PCR products were submitted to capillary electrophoresis. In the blood donors, the ranges of the five *loci* were: SCA1, 19 to 36 (CAG)n; SCA2, 6 to 28 (CAG)n; SCA3, 12 to 34 (CAG)n; SCA6, 2 to 13 (CAG)n; and SCA7, 2 to 10 (CAG)n. No deviations from Hardy-Weinberg equilibrium were detected. In the ataxia group, we found (CAG)n above the range of the asymptomatic blood donors in SCA3 (21.74%) followed by SCA2 (5.22%), SCA7 (2.61%), SCA6 (0.87%), and no cases of SCA1. The remaining 80 cases (69.56%) have different diagnoses from the type here studied. These data defined the alleles and their frequencies, as well as demonstrated their stability in the population not affected. The molecular diagnosis test confirmed the clinical diagnosis in 28/45 cases and classified another 7/70 from the clinical unclassified ataxias group.

KEY WORDS: microsatellites repeats, allelic repeats, spinocerebelar ataxia, capillary electrophoresis, population genetics, (CAG)n in Brazilians.

Ataxias espinocerebelares: freqüência de alelos e microsatélites em indivíduos normais e afetados

Resumo – A incidência e o diagnóstico das ataxias espinocerebelares (SCA) é algumas vezes difícil de avaliar devido a sobreposição dos diversos subtipos e por algumas serem devido a mutações das expansões do mesmo trinucleotídeo CAG. Para investigar a incidências das SCA no sul do Brasil, analisamos as repetições do trinucleotídeo (CAG)n nos *loci* das SCA1, SCA2, SCA3, SCA6 e SCA7, a fim de identificar os seus limites e freqüência. Examinamos o sangue de 154 doadores de sangue assintomáticos e 115 pacientes com ataxias progressivas. O produto do PCR do sangue foi submetido a eletroforese capilar. Nos doadores de sangue, as expansões encontradas nos cinco *loci* foram: SCA1, 19 a 36 (CAG)n; SCA2, 6 a 28 (CAG)n; SCA3, 12 a 34 (CAG)n; SCA6, 2 a 13 (CAG)n; and SCA7, 2 a 10 (CAG)n. Não foi detectado desequilíbrio na equação de Hardy-Weinberg. No grupo das ataxias encontramos repetições CAG acima das freqüências dos doadores de sangue na SCA3 (21,7%), seguido da SCA2 (5,22%), SCA7 (2,61%), SCA6 (0,8%) e não foi encontrado nenhum caso de SCA1. Os 80 casos restantes (69,56%) devem ter uma forma de ataxia diferente das aqui estudadas. Esses dados definem os alelos e suas freqüências, bem como demonstram a sua estabilidade na população não afetada. O diagnóstico molecular confirmou o diagnóstico clínico em 28/45 dos casos e permitiu classificar outros 7/70 que pertenciam ao grupo de ataxias clinicamente não classificadas.

PALAVRAS- CHAVE: repetição de microsatélites, repetição de alelos, ataxia espinocerebelar, eletroforese capilar, genética de populações, (CAG)n em Brasileiros.

Neuromuscular and Molecular Biology Unit, Neurology Division, Internal Medicine Department, Hospital de Clínicas, Universidade Federal do Paraná, Curitiba PR, Brazil.

This study was funded by a grant from Fundação Araucária, Curitiba, Brazil, and partially supported by CNPq (Conselho Nacional de Pesquisas) and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), Brasília, Brazil.

Received 28 July 2009, received in final form 30 July 2009. Accepted 14 August 2009.

Dr. Lineu César Werneck – Neuromuscular and Molecular Biology Unit / Neurology Division / Internal Medicine Department / Hospital de Clínicas / Universidade Federal do Paraná - Rua General Carneiro 181 / 310 - 80060-900 Curitiba PR - Brasil. E-mail: werneck@ufpr.br

Spinocerebellar ataxia (SCA) subtypes 1, 2, 3 (MJD), 6 (CACNA1), and 7 are late-onset disorders with an autosomal dominant pattern. SCAs are a heterogeneous group of neurodegenerative disorders with progressive deterioration in balance and coordination. The clinical diagnosis of each one of the several types is complicated by the overlap of the phenotypes between genetic subtypes. Prevalence of SCA is estimated to be 3 in 100,000 people. However, it is possible that this figure is underestimated. The prevalence of specific subtypes depends on the ethnic origin and geographic location of the populations studied[1]. The mutations associated with the SCA1, SCA2, SCA3, SCA6 and SCA7 (chromosomes 6p23, 12q23-24.1, 14q21, 19p and 3p21.2-12, respectively) subtypes are caused by CAG trinucleotide repeat expansions in the patient's genes. These expansions encode polyglutamine repeats. The number of CAG trinucleotide repeats varies from one population to another[2-4]. In normal individuals, SCA 1 alleles range from 6 to 39 CAG repeats (CAGn)[5-8]; in SCA2 alleles range from 15 to 24 CAGn[9-11]; in SCA3 (Machado Joseph disease – MJD) from 13 to 41 CAGn[12-14]; in SCA6¨from 6 to 16 CAGn[15-16], and in SCA7 ranges from 4 to 17 CAGn[17-18]. Intermediate alleles are found in studies with families which have not been described in the normal population. Individuals that present intermediate alleles between normal and affected may not clinically manifest the disease, while other individuals with the same allele may show signs and symptoms, and can be considered to have alleles with reduced penetrance. The presence of these three distinct groups of alleles shows the importance of defining the range of CAG repetitions for normal individuals and patients for each population[19].

Population genetic analysis of expansion diseases suggests a controversial pathogenic mechanism among them[20-23]. There is the hypothesis that most cases of new mutations that cause expansion diseases arise from the existence of a large number of normal repetitions or intermediate alleles in the non-affected population. These alleles have been considered reservoirs for the generation of new expanded alleles[24]. The ability to correlate the prevalence of a particular expansion disease with the frequency of highest size of normal alleles in the respective gene depends directly on the epidemiological profile of these diseases. This correlation was no observed among Portuguese with MJD and no intermediate alleles were seen in the normal population[23]. Moreover, it is believed that expanded alleles in MJD are maintained in the population over generations by presenting greater adaptive value[25]. Allelic frequencies in the normal population could explain the evolutionary mechanism to support expanded alleles through generations[20,23].

The present study is an analysis of trinucleotide (CAG) n repeats in unaffected and affected individuals from Southern Brazil (Paraná) of SCA1, SCA2, SCA3, SCA6 or SCA7 ataxia in order to define protocols, allelic frequency and repeat number ranges (number of CAG repeats).

METHOD

DNA samples

The sample consisted of 154 asymptomatic blood donors from Southern Brazil (Paraná) chosen at random and 115 patients with progressive ataxia after being ruled out structural disorder of the central nervous system. The suspicion diagnosis according the clinical findings were SCA2 (7 cases), SCA3 (30 cases), SCA4 (1 case), SCA6 (5 cases), SCA7 (3 cases), SCA8 (1 case), SCA10 (7 cases), Friedreich ataxia (2 cases) and unclassified ataxias by clinical findings (59 cases). None of the blood donors included in this sample had a clinical diagnosis or family history of spinocerebellar ataxia. The donor group consisted of 90 males and 64 females, with a mean age of 32.29±10.63 years. The population breakdown of the blood donor group was as follows: 72.46% Euro-Brazilians, 2.17 % Afro-Brazilians, 23.19 % admixed and 2.17% Amerindians.

Blood samples were collected in tubes with EDTA (BD Vacutainer), and genomic DNA was extracted from peripheral leukocytes using a phenol-chloroform procedure as previously described[26]. Written informed consent was obtained from each individual. The study had been previously approved by the Ethics Committee of the Hospital de Clínicas – Universidade Federal do Paraná.

PCR amplification and capillary electrophoresis

The five loci were amplified in three different reactions with 10 to 30 ng of genomic DNA using an MJ Research model PTC thermal cycler (MJ Research, inc., Watertown, MA, USA). Primers were end-labeled with fluorescent dyes appropriate for ABI PRISM instruments (Applied Biosystems, Foster City, CA, USA). Fluorescent labels were selected based on the normal allele size ranges described above for each locus, while taking care to ensure that loci with overlapping size ranges were given different labels. The following reactions were carried out: one triplex reaction with 1.5 pmol of each primer for SCA1 (36), SCA2 (10) and SCA6 (18); a single reaction with 12pmol of each SCA3 primer (SCA3F 5'NED-CTGGCCTTTCACATGG 3' and SCA3R 5' CCAGTCACTACTTTGATTCGTG 3') and a reaction with 15 pmol of each SCA7 primer (SCA7F 5'VIC- TTGTAGGAGCGGAAAGAATGTC 3'and SCA7R 5' CTTCAGGACTGGGCAGAGG 3'). The last two pairs of primers (SCA3 pair and SCA7 pair) were designed for the present study. The triplex reaction was performed with AccuPrime DNA polymerase systems (1X buffer BII – 2mM of each dNTPs; 1.5 mM of MgCl2; 200mM Tris-HCl pH 8.4 and 1.5U AccuPrime *Taq* DNA polymerase) (Invitrogen/Stratagene, La Jolla, CA, USA). Both the SCA3 and SCA7 reactions were performed with 200 mM of each dNTP (Amershan Pharmacia Inc., Piscataway, NJ, USA), 1X reaction buffer for SCA3 and 0.7X for SCA7 (10mM Tris-HCl pH9.0, 50mM KCl and 0.1% Triton X-10); 1.5

mmol/L MgCl2 and 1.5 U YT1 *Taq* DNA polymerase (Promega, Madison, WI, USA). In the SCA7 reaction, 4% of DMSO was added. All reactions had a final volume of 15 μL. The triplex reaction conditions were 1 cycle at 96°C for 10 min; 32 cycles at 94°C for 1 min, a ramp rate of 0.2°C/s, 61°C for 1 min 30sec and 68°C for 1 min 30 sec and 1 cycle at 68°C for 1 hour. The SCA3 reaction conditions were 1 cycle at 95°C for 10 min; 5 cycles at 95°C for 1 min, 59°C for 1 min and 72°C for 1 min 30 sec; 20 cycles at 95°C for 1 min, 60°C for 1 min and 72°C 1min 30 sec; 7 cycles at 95°C for 1 min, 61°C for 1 min and 72°C for 1 min 30 sec; and a final extension at 72°C for 1 hour. The SCA7 reaction conditions were 1 cycle at 96°C for 10 min; 5 cycles at 95°C for 1 min, 51°C for 1 min and 72°C for 1 min 30 sec; 21 cycles at 95°C for 1 min, 52°C for 1 min and 72°C for 1 min 30 sec; 7 cycles at 95°C for 1 min, 55°C for 1 min and 72°C for 1min 30 sec; and a final extension at 72°C for 1 hour. Samples were prepared by adding 1 μL of each PCR reaction to 5 μL of deionized formamide and 0.25 μl of internal molecular weight standard (GS-500 ROX). After denaturizing, the samples were injected into an ABI PRISM 3100-Avant Genetic analyzer with a 36 cm × 50 μm capillary containing Performance Optimized Polymer-4 (POP-4) for 5 sec. Amplicon length was calculated by comparison with the molecular weight standard using the Genemapper program.

The sizes of the PCR products were calculated automatically on the basis of a standard curve from the internal size standard. Each allele represents the number of CAG repeats. While different individuals had the same alleles, these differed slightly in size (bp) from the theoretical values for amplicon length of trinucleotide-repeat region. To define the alleles reliably, those individuals with alleles whose sizes were close to the theoretical values were grouped together. Using statistical analysis, the median and variance for clinically normal patients was determined. Those values are used in Genemapper program to determinate the alleles of each loci.

Statistical and genetic analysis

The number of CAG repeats was calculated and statistical analysis performed using SPSS software. Hardy-Weinberg (HW) proportions and heterozygosis (H) were calculated using Genepop software[27,28].

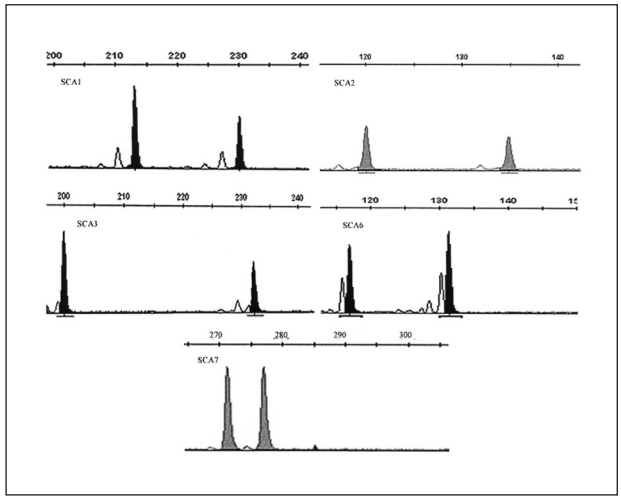

Fig 1. Amplification of normal-range alleles at each locus analyzed - well-defined peaks with no background. SCA1: Spinocerebellar Ataxia type 1; SCA2: Spinocerebellar Ataxia type 2; SCA3: Spinocerebellar Ataxia type 3; SCA6: Spinocerebellar Ataxia type 6; SCA7: Spinocerebellar Ataxia type 7.

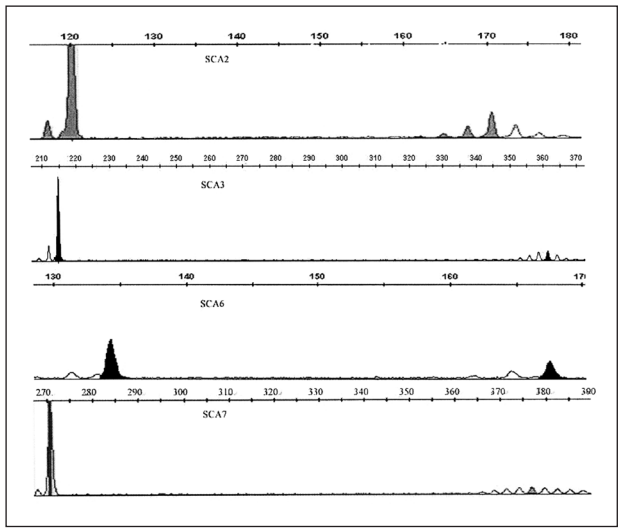

Fig 2. Amplification of expanded alleles from patients with spinocerebellar ataxia. SCA2: Spinocerebellar Ataxia type 2; SCA3: Spinocerebellar Ataxia type 3; SCA6: Spinocerebellar Ataxia type 6; SCA7: Spinocerebellar Ataxia type 7.

RESULTS

The SCA1, SCA2, SCA3, SCA6 and SCA7 loci were analyzed in 308 chromosomes for each locus from 154 normal individuals from Southern Brazil and in 230 chromosomes for each locus from 115 individuals with clinical suspicion of ataxia. The PCR conditions resulted in consistent amplification of the normal-size alleles analyzed, which are seen as well-defined peaks with no background and abnormal expanded are seen with stutter band from somatic mosaicism and technical artifacts (Figs 1 and 2).

Inter-assay variability of the same allele among different individuals at each SCA locus (calculated as a 95% confidence interval) ranged from 0.11 to 1.66 bp. For the most frequent alleles at each locus, a higher variance (1.10) was observed in allele 5 of SCA7. The intra-assay variability was taken repeating five times the same product of amplification of five different individuals. This was carried out for the five loci. This variability (calculated as a 95% confidence interval) also was between 0.02 to 0.69 bp.

The allelic frequency for SCA1 *locus* the allele size ranged from 19 to 36 (CAG)n with 16 different alleles (Table 1). The SCA2 locus allele size ranged from 6 to 28 (CAG)n with 13 different alleles (Table 2). The allele sizes

Table 1. Allelic frequencies (±SE%) at the five loci analyzed for spinocerebellar ataxia 1 (SCA1) among controls and patients.

SCA 1		
Alleles	Controls	Patients
*19	0.35±0.34	–
*20	0.35±0.34	–
*23	0.35±0.34	0.54±0.54
*24	0.35±0.34	–
*25	0.35±0.34	0.54±0.54
*26	2.78±0.94	2.15±1.06
*27	2.78±0.94	2.15±1.06
*28	7.29±1.48	4.84±1.57
*29	28.11±2.56	27.42±3.27
30	32.29±2.66	37.1±3.54
31	10.76±1.77	7.53±1.93
32	7.99±1.54	11.83±2.37
33	2.78±0.94	3.76±1.39
34	1.39±0.67	–
35	1.04±0.58	0.54±0.54
36	1.04±0.58	1.6±0.92

SE: Standard error of mean; –: not detected; *: alleles not described in other Brazilian populations.

Table 2. Allelic frequencies (±SE%) at the five loci analyzed for spinocerebellar ataxia 2 (SCA2) among controls and patients.

SCA 2		
Alleles	Controls	Patients
**6	0.35±0.34	–
*14	0.35±0.34	0.43±0.48
*15	0.35±0.34	0.43±0.48
*17	0.35±0.34	2.17±1.06
19	0.35±0.34	0.87±0.67
20	84.44ff11.95	84.35±2.64
21	7.29±1.48	7.00±1.85
22	1.04±0.57	0.43±0.48
23	0.69±0.47	–
*25	1.74±0.74	0.87±0.67
*26	0.35±0.34	–
*27	0.35±0.34	–
*28	0.35±0.34	0.87±0.67
37	–	**0.43±0.48**
38	–	**1.30±0.82**
43	–	**0.43±0.48**
56	–	**0.43±0.48**

SE: Standard error of mean; –: not detected; *: alleles not described in others Brazilian populations; **: alleles not described in other populations. Expanded alleles are in bold case.

Table 3. Allelic frequencies (±SE%) at the five loci analyzed for spinocerebellar ataxia 3 (SCA3) among controls and patients.

SCA 3		
Alleles	Controls	Patients
12	26.57±2.53	20.43±2.89
13	0.35±0.34	1.74±0.93
14	–	0.43±0.48
15	1.4±0.67	0.43±0.48
16	0.7±0.48	3.04±1.23
17	2.45±0.88	4.34±1.46
18	10.84±1.78	11.3±2.27
19	3.85±1.10	5.22±1.60
20	27.96±2.57	16.52±2.67
21	5.94±1.35	7.39±1.88
22	1.4±0.67	3.04±1.23
23	1.4±0.67	3.47±1.32
24	12.24±1.87	6.09±1.72
25	2.1±0.82	2.61±1.14
26	0.7±0.48	2.17±1.05
27	0.35±0.34	0.43±0.48
28	0.7±0.48	–
29	0.35±0.34	0.43±0.48
34	0.7±0.48	–
62	–	**0.87±0.67**
64	–	**0.87±0.67**
65	–	**1.74±0.94**
66	–	**1.30±0.81**
67	–	**0.87±0.67**
68	–	**1.74±0.94**
69	–	**1.30±0.81**
70	–	**1.03±0.72**
71	–	**0.43±0.48**
72	–	**0.43±0.48**
74	–	**0.43±0.48**

SE: Standard error of mean; –: not detected; *: alleles not described in other Brazilian populations. Expanded alleles are in bold case.

for the SCA3 locus ranged from 12 to 34 (CAG)n with 18 different alleles dispersed through the locus (Table 3). In the blood donors, the SCA6 allele size ranged from 2 to 13 (CAG)n with 8 different alleles (Table 4). The SCA7 allele size ranged from 2 to 10 (CAG)n with 7 different alleles (Table 5). No expanded or intermediate alleles were found at SCA1, SCA2, SCA3, SCA6 and SCA7 *loci*. Among patients the number of alleles ranged from 23 to 36 (CAG) n in SCA1 with 12 different alleles. No expanded alleles were found in this locus (Table 1). In locus SCA2 the size

Table 4. Allelic frequencies (±SE%) at the five loci analyzed for spinocerebellar ataxia 6 (SCA6) among controls and patients.

SCA 6		
Alleles	Controls	Patients
**2	2.08±0.81	2.66±1.17
4	7.98±1.54	1.06±0.75
5	0.35±0.34	5.32±1.64
8	39.59±2.79	35.11±3.48
9	18.4±2.21	14.36±2.56
10	29.17±2.59	33.52±3.44
11	2.08±0.81	6.91±1.85
12	–	0.53±0.53
13	0.35±0.34	–
21	–	**0.53±0.53**

SE: Standard error of mean; –: not detected; **: alleles not described in other populations. Expanded alleles are in bold case.

Table 5. Allelic frequencies (±SE%) at the five loci analyzed for spinocerebellar ataxia 7 (SCA7) among controls and patients.

SCA 7		
Alleles	Controls	Patients
**2	1.15±0.67	2.17±1.34
4	70.61±2.88	64.38±4.38
5	8.4±1.75	10.00±2.75
6	16.03±2.32	18.26±3.54
7	2.67±1.02	3.48±1.68
8	0.76±0.55	–
10	0.38±0.39	–
39	–	**0.87±0.85**
50	–	**0.43±0.60**

SE: Standard error of mean; –: not detected; **: alleles not described in other populations. Expanded alleles are in bold case.

Table 6. Relationship between clinical and molecular diagnosis in 115 patients with progressive ataxia.

	Molecular diagnosis		
	Positive	Negative	
Clinical diagnosis (SCA2, SCA3, SCA6, SCA7)	28 (a)	17 (b)	45 (a+b)
Unclassified progressive ataxias	7 (c)	63 (d)	70 (c+d)
	35 (a+c)	80 (b+d)	115
	χ^2=32.86	p=0.000	

Sensibility: a/(a+c)=80.00 % VPP: a/(a+b)=62.22 %
Specificity: d/(b+d)=78.75 % VPN: d/(c+d)=90.00 %

Table 7. Clinical and molecular diagnosis in 115 cases of progressive ataxias.

Clinical diagnosis		Molecular diagnosis					
	Number of cases	SCA1	SCA2	SCA3	SCA6	SCA7	Total positives
SCA1	–	–	–	–	–	–	–
SCA2	7	–	5	–	–	–	5
SCA3	30	–	–	20	–	–	20
SCA6	5	–	–	–	1	–	1
SCA7	3	–	–	–	–	2	2
Clinically unclassified ataxias	70	–	1	5	–	1	7
Total	115	–	6	25	1	3	35

–: Absent. Cases confirmed only by molecular diagnosis are in bold.

of the alleles ranged from 14 to 56 (CAG)n and four abnormal expanded alleles were found in six patients (Table 2). The locus SCA3 have the size of the alleles ranging from 12 to 74 (CAG)n with 28 different alleles. There are 11 expanded alleles in 25 patients (Table 3). The size between alleles at SCA6 ranged from 2 to 21 (CAG)n with 9 different alleles. Only one expanded allele was found at the SCA6 locus in only one patient (Table 4). These patients at SCA7 locus showed the size of the alleles ranging from 2 to 50 (CAG)n with two expanded alleles in three patients (Table5).

Among the 115 patients with progressive cerebellar ataxia tested, the molecular diagnosis was positive in 35 cases (Table 6).

Table 8. Frequency of diseases at 115 individuals with clinical suspicion of ataxia.

Diseases	Frequency (%)	Number of individuals
SCA1	0	0
SCA2	5.22	6
SCA3	21.74	25
SCA6	0.87	1
SCA7	2.61	3
Clinically unclassified ataxias	69.56	80

The molecular tests confirmed the clinical suspicion in 5 of 7 cases of SCA2, in 20 of 30 cases of SCA3, in 1 of 5 cases of SCA6 and in 2 of 3 SCA7. In the group of unclassified ataxias, was found 1 cases SCA2, 5 cases of SCA3 and 1 case of SCA7 (Table 7).

At these five subtypes of ataxia, the most frequent among patients is SCA3, followed by SCA2, SCA7 and SCA6. No case of SCA1 was found (Table 8).

All individuals with expansion featuring are heterozygous, so they showed a normal allele and the other allele expanded.

No deviations from Hardy-Weinberg (HW) equilibrium were detected and the heterozygosis (H) data for the five loci are in Table 9. Significant differences in mean allelic size (number base pairs) between patients and control samples were found for SCA2 and SCA3 *loci* (Table 10).

DISCUSSION

Capillary electrophoresis (CE) has been used for DNA fragment analysis in a number of studies to determine the number of microsatellite repeats in expansion illnesses[29-31].

In the present study, small variations in allele size were observed between individuals, although the alleles were close to the expected theoretical sizes. This variation is probably due to factors such as electrophoresis temperature, the polymer batch number, and capillary age[29-31].

Methods to define alleles in these SCA loci by comparing the results of capillary electrophoresis (CE) with those of PAGE assay have been published and show a discrepancy in amplicons sizes between these methodologies. In CE, amplicons were found to migrate faster than predicted, and the results show a lower number of repeats in one or two peaks[29-31]. Factors contributing to these discrepancies could be the different sieving effects of cross-linked PAGE compared with the linear polymer used in CE, the electroosmotic effect, the extent of denaturation or differences in secondary structure between amplicons and standard markers[32]. Considering the previous studies regarding differences between PAGE with CE, we did not compare these two methodologies. Since the standard deviation for each allele was small the use of GS-500 ROX provided reliable results. To estimate the range of each allele we analyzed the samples of normal asymptomatic population and defined a pool of existing alleles in these individuals. When the mean allelic size of patients was compared with normal asymptomatic population we found a significant difference in the SCA2 and SCA3 (Table 10).

Table 9. Heterozygosis and Hardy-Weinberg equilibrium (Qui Square (χ^2), probability and degrees of freedom) in five loci analyzed.

Loci	χ^2	p	Degrees of freedom	Heterozygosis (H)
SCA1	12.87	0.80	18	0.79
SCA2	2.10	0.84	5	0.24
SCA3	20.13	0.33	18	0.82
SCA6	4.96	0.97	12	0.72
SCA7	12.87	0.80	7	0.47

Table 10. Comparison of all alleles from 115 patients and control asymptomatic population ("t" test for independent samples) in spinocerebellar ataxias (SCA).

	M_c	M_p	N_c	N_p	SD_c	SD_p	t	p
$SCA1_C$ vs. $SCA1_P$	29.81	29.97	308	230	2.02	1.68	–0.94	0.35
$SCA2_C$ vs. $SCA2_P$	20.18	20.84	308	230	1.35	4.13	–2.26	**0.025**
$SCA3_C$ vs. $SCA3_P$	18.48	24.75	308	230	4.5	17.22	–5.23	**0.00**
$SCA6_C$ vs. $SCA6_P$	8.51	8.89	308	230	1.68	2.64	–1.87	0.062
$SCA7_C$ vs. $SCA7_P$	4.5	5.2	308	230	0.99	5.12	–1.79	0.076

c: controls; p: patients; M: mean; N: number of chromossomes; SD: standard deviation; p: probability. Significant values in bold case.

The 36-repeat allele in the SCA1 locus is borderline in some populations being an intermediate allele with reduced penetrance[8]. Therefore, patients who have this allele should be considered negative for SCA1 in southern Brazilian population. Our data found no intermediate alleles for SCA2, SCA3, SCA6 or SCA7 in the unaffected population. However, studies of families with SCA traits usually reveal an intermediate allele size[3,8,10,33]. Probably, intermediate alleles are as rare as the expanded alleles and it must be found only in affected families with the disease.

A 22-repeat allele is described in other populations as the most frequent at the SCA2 *locus*[10,11,24]. Molecular analysis of unrelated Brazilian families with SCA3 showed that normal alleles ranged from 12 to 33 (CAG)n[34], and in other populations studied the two major alleles had 14 and 23 (CAG)n[13,35]. Comparisons of allelic frequencies must be made using population studies based on the CE methodology. Despite the differences in allele size ranges found in this study could be attributed to methodological factors, discrepancies in alleles size ranges are important to characterize the Brazilian population since each human population could present their own genetic profile different from others.

Intermediate allele values tend to trigger the growth of repetitions leading to the expansion. There are studies of clinically normal individuals with MJD in different populations that support the idea that the presence of alleles of size greater than or equal to 27 (CAG)n correlates with higher prevalence of MJD in patients[35]. According to this study we found alleles with a size equal to or greater than 27 (CAG)n in the MJD gene, and MJD is the most frequent ataxia among patients. These findings have an epidemiological interest as these alleles appear to be in haplotypes that have a great potential for expansion by gene conversion[19,36-40].

Alleles with a number of (CAG)n greater than 27 are observed in the Portuguese, American, Japanese and African populations, of which there is a high incidence in SCA3 patients[13,37,41]. In the patient group in the present study, SCA3 locus had more expanded alleles than other *loci*. This corroborates the hypothesis that the Portuguese and African populations brought a pool of founding genes to Brazil with a common mutation in the SCA3 gene identified as the cause of MJD (SCA3)[37,41-43]. However, other studies of the Portuguese population did not show any correlation between the frequency of the large normal alleles and the frequency of the disease[20-23]. There was no significant correlation between the frequency of MJD and the frequency of small, medium or large normal alleles in the SCA3 locus of the Portuguese population[23]. Most of the mutant chromosomes worldwide share a common haplotype, and a second haplotype were found to be present in a group of Portuguese patients. Two other additional haplotypes were found in some families. These haplotypes may have derived from a main haplotype as a result of microsatellite mutations[44]. Studies associating intragenic haplotypes with intergenerational repeat length instability in the normal chromosome did not support the hypothesis of pre-mutation[40]. The results of the present study, in which no expanded or intermediate alleles were found in the normal population, corroborate the studies cited above.

Even genetic *loci* with very similar molecular characteristics have different expansion dynamics. The widespread distribution of the wild-type MJD alleles is attributed to a high mutation rate, without any bias to expansion, which could explain the observed allelic frequencies. This is in accordance with a model that postulates the presence of a small number of mutational events from which most of the MJD cases worldwide have originated[20] and contrasts with findings for Huntington disease (HD)[21], in which intermediate and expanded alleles are observed in the normal population. The results of these studies may corroborate the idea that there is a mechanism that maintains these diseases in the population and suggest the existence of a few founding chromosomes from which all families originated.

No correlation could be made in the locus SCA 3 only with the data found in this work because patients with MJD is most common form of ataxia and in the sample control no alleles intermediate 46 to 56 (CAG)n) were observed in normal alleles and 2.1% of alleles size range from 27 up to 34 (CAG)n. A study with a larger sample, in which there could be a stratification by Brazilian regions with distinct frequencies of MJD, for example, could lead to a positive or negative correlation between the highest incidence of the disease with normal alleles whose number of (CAG)n are larger than 27. Furthermore, the population study of alleles with potential to expansion can lead to accurate understanding of how alleles are stable avoiding the appearance of alleles expanded. Another interesting fact is that intermediate alleles have ambiguity between the possibility of normal and patient phenotype and is important elucidate possible cis elements involved with molecular mechanisms that may lead to one or the other phenotype.

The present study provides subsidy to understand the origin of the mutations in loci SCA 1, SCA 2, SCA 3, SCA 6 and SCA 7. However, a haplotypic study including other sites of these loci should be carried out in both samples. This is an initial study and in the future other types of ataxia due to CAG repetitions will be done, such as SCA 12, SCA17 and DRPL, to expand the molecular diagnosis at our center.

The unaffected population is in HW equilibrium for the five *loci* analyzed that demonstrated allelic stabili-

ty in this population. Data of heterozygosis (H) characterized the population as the amount heterozygous for each gene. The heterozygosis in this study for the five loci is similar to that that found elsewhere[45]. Heterozygosis similar to the locus ATXN 3 is described in other populations of the world (H=0.84, H=0.86)[37]. This result combined with no deviation of HW equilibrium indicates that the region of (CAG)n in the gene of MJD can be used as a marker in studies of population genetics[23].

The frequency of subtypes of SCA 1, 3 and 6 found in this study is similar to those found in others reports but different for SCA2 and SCA7[34,45,46]. The differences between these results may be due to distinct loci analyzed and probably because the sample of Brazilian patients come from different regions of the country. These results are in agreement with the postulations by different authors on the variability of alleles and frequency of subtypes of ataxia in different populations. Furthermore, this difference emphasizes the importance of population study in different regions of Brazil due to the ethnic diversification in the constitution of the country.

In conclusion, the statistical data from our study allowed us to identify the alleles and their frequencies. This study defined the normal range for each one of the five subtypes and found expanded alleles in patients. Population genetic analysis ensured the technical quality of this assay and allowed the normal range of (CAG)n size repeats to be defined in order to distinguish between unaffected and affected individuals for diagnostic purposes. To understand the origins of the mutations at these loci it is important to consider the information provided by the normal range variation so as to understand the dynamics of these loci and provide data to be useful from Southern Brazil.

REFERENCES

1. Schöls L, Bauer P, Schmidt T, Schulte T, Riess O. Autosomal dominant cerebellar ataxias: clinical features, genetics, and pathogenesis. Lancet Neurol 2004;3:291-304.
2. Pujana MA, Corral J, Gratacòs M, et al. Spinocerebellar ataxias in Spanish patients: genetic analysis of familial and sporadic cases: the Ataxia Study Group. Hum Genet 1995;104:516-522.
3. Ranum LP, Lundgren JK, Schut LJ. Spinocerebellar ataxia type1 and Machado-Joseph disease:incidence of CAG expansions among adult-onset ataxia patients from 311 families with dominant, recessive, or sporadic ataxia. Am J Hum Genet 1995;57:603-608.
4. van de Warrenburg BP, Hendriks H, Dürr A, et al. Age at onset variance analysis in spinocerebellar ataxias: a study in a Dutch-French cohort. Ann Neurol 2005;57:505-512.
5. Genis D, Matilla T, Volpini V, et al. Clinical, neuropathologic, and genetic studies of a large spinocerebellar ataxia type 1 (SCA1) kindred: (CAG)n expansion and early premonitory signs and symptoms. Neurology 1995;45:24-30.
6. Matilla T, Volpini V, Genis D, et al. Presymptomatic analysis of spinocerebellar ataxia type 1 (SCA1) via the expansion of the SCA1 CAG-repeat in a large pedigree displaying anticipation and parental male bias. Hum Mol Genet 1993;2:2123-2128.
7. Suzuki Y, Sasaki H, Wakisaka A, et al. Spinocerebellar ataxia 1 (SCA1) in the Japanese: analysis of CAG trinucleitide (sic) repeat expansion and instability of the repeat for paternal transmission. Jpn J Hum Genet 1995;40:131-143.
8. Zuhlke C, Dalski A, Hellenbroich Y, Bubel, S, Schwinger, Burk K. Spinocerebellar ataxia type 1 (SCA1): phenotype-genotype correlation studies in intermediate alleles. Europ J Hum Genet 2002;10:204-209.
9. Cancel G, Durr A, Didierjean O, et al. Molecular and clinical correlations in spinocerebellar ataxia 2: a study of 32 families. Hum Mol Genet 1997;6:709-715.
10. Pulst SM, Nechiporuk A, Nechiporuk T, et al. Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. Nat Genet 1996;14:269-276.
11. 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. Nat Genet 1996;14:277-284.
12. Giunti P, Sweeney MG, Harding AE. Detection of the Machado-Joseph disease/spinocerebellar ataxia three trinucleotide repeat expansion in families with autosomal dominant motor disorders, including the Drew family of Walworth. Brain 1995;118:1077-1085.
13. Kawaguchi Y, Okamoto T, Taniwaki M. CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. Nat Genet 1994; 8:221-228.
14. van Alfen N, Sinke RJ, Zwarts MJ, et al. Intermediate CAG repeat lenghts (53, 54) for MJD/SCA 3 are associated with an abnormal phenotype. Ann Neurol 2001;49:805-807.
15. Matsuyama Z, Kawakami H, Maruyama H, et al. Molecular features of the CAG repeats of spinocerebellar ataxia 6 (SCA6). Hum Mol Genet 1997;6:1283-1287.
16. Zhuchenko O, Bailey J, Bonnen P. Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha 1A-voltage-dependent calcium channel. Nat Genet 1996;15:62-69.
17. 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.
18. Stevanin G, Giunti P, Belal GD, et al. De novo expansion of intermediate alleles in spinocerebellar ataxia 7. Hum Molec Genet 1998;7:1809-1813.
19. Andrew SE, Goldberg YP, Hayden MR. Rethinking genotype and phenotype correlations in polyglutamine expansion disorders. Hum Mol Genet 1997;6:2005-2010.
20. Andres A, Lao O, Soldevila M, Calafell F, Bertranpetit J. Dynamics of CAG repeat loci revealed by the analysis of their variability. Hum Mutat 2002;21:61-70.
21. Costa MC, Guimarães L, Ferreirinha F, Sousa A, Maciel P, Sequeiros J. Study of the normal CAG tract at the Huntington disease locus in the Portuguese population. Eur J Hum Genet 2002;10:263-264.
22. Deka R, Miki T, Yin SJ, et al. Normal CAG repeat variation at the DRPLA locus in world populations. Am J Hum Genet 1995;57:508-511.
23. Lima M, Costa MC, Montiel R, et al. Population genetics of wild-type CAG repeats in the Machado-Joseph disease gene in Portugal. Hum Hered 2005;60:156-163.
24. 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;14:285-291.
25. Prestes PR, Saraiva-Pereira ML, Silveira I, Sequeiros J, Jardim LB. Machado-Joseph disease enhances genetic fitness: a comparison between affected and unaffected women and between MJD and the general population. Ann Hum Genet 2008;72:57-64.
26. Sambrook J, Fritsch EF, Maniats T. Molecular Cloning, 2nd Ed. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory, 1989:16-19.
27. Hartl D. Génétique des Population, Edition Française. Editora Médicine-Sciences Flammarion, Paris, 1998:5-34.
28. Raymond M, Rousset F. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicist. J Hered 1995;86:248-249.
29. Dorschner MO, Barden D, Stephens K. Diagnosis of five spinocerebellar ataxia disorders by multiplex amplification and capillary electrophoresis. J Mol Diagn 2002;4:108-113.
30. Larsen LA, Gronskov K, Norgaard-Pedersen B, Brondum-Nielsen K, Hasholt L, Vuust J. High-throughput analysis of fragile X (CGG)n alleles in the normal and premutation range by PCR amplification and automated capillary electrophoresis. Hum Genet 1997;100:564-568.

31. Le H, Fung D, Trent RJ. Applications of capillary electrophoresis in DNA mutation analysis of genetic disorders. Mol Pathol 1997;50:261-265.
32. Rosenblum BB, Oaks F, Menchen S, Johnson B. Improved single-strand DNA sizing accuracy in capillary electrophoresis. Nucleic Acids Res 1997;25:3925-3929.
33. Takiyama Y, Oyanagi S, Kawashima S, et al. A clinical and pathologic study of a large Japanese family with Machado-Joseph disease tightly linked to the DNA markers on chromosome 14q. Neurology 1994;44: 1302-1308.
34. Lopes-Cendes I, Teive HGA, Cardoso F et al. Molecular characteristics of Machado-Joseph disease mutation in 25 newly described Brazilian families. Braz J Genet 1997;20:717-724.
35. Takano H, Cancel G, Ikeuchi T. Close associations between prevalence of dominantly inherited spinocerebellar ataxias with CAG-repeat expansions and frequencies of large normal CAG alleles in Japanese and Caucasian populations. Am J Hum Genet 1998;63:1060-1066.
36. Bauer PO, Zumrova A, Matoska V, et al. Absence of spinocerebellar ataxia type3/Machado-Joseph disease within ataxia patients in the Czech population. Eur J Neurol 2005;12:851-857.
37. Limprasert P, Nouri N , Heyman RA, et al. Analysis of CAG repeat of the Machado-Joseph gene in human, chimpanzee and monkey populations: a variant nucleotide is associated with the number of CAG repeats. Hum Mol Genet 1996;5:207-213.
38. Chattopadhyay B, Basu P, Gangopadhyay PK. Variation of CAG repeats and two intragenic polymorphisms at SCA3 locus among Machado-Joseph disease/SCA3 patients and diverse normal populations from eastern India. Acta Neurol Scand 2003;108:407-414.
39. Mittal U, Srivastava AK, Jain S, Mukerji M. Founder haplotype for Machado-Joseph disease in the Indian population. Arch Neurol 2005; 62:637-640.
40. Maciel P, Gaspar C, Guimaraes L. Study of three intragenic polymorphisms in the Machado-Joseph disease gene (MJD1) in relation to genetic instability of the (CAG)n tract. Eur J Hum Genet 1999;7:147-156.
41. Teive HAG, Arruda WO. A família Drew de Walworth - um século após a avaliação inicial finalmente o diagnóstico. Arq Neuropsiquiatr 2004; 62:177-180.
42. Silveira I, Lopes-Cendes I, Kish S, et al. Frequency of spinocerebellar ataxia type1, dentatorubropallidoluysiana atrophy and Machado Joseph disease mutations in a large group of spinocerebellat ataxia patients. Neurology 1996;46:214-218.
43. Harding AE. Clinical features and classification of inherited ataxias. Adv Neurol 1993;61:1-14.
44. Gaspar C, Lopes-Cendes I, Hayes S, et al. Ancestral origins of the Machado-Joseph disease mutation: a worldwide haplotype study. Am J Hum Genet 2001;68;523-528.
45. Silveira I, Miranda C, Guimarães L, et al. Trinucleotide repeats in 202 families with ataxia: a small expanded (CAG)n allele at the SCA 17 *locus*. Arch Neurol 2002;59:623-629.
46. Jardim LB, Silveira I, Pereira ML. A survey of spinocerebellar ataxia in south Brazil - 66 new cases with Machado-Joseph disease, SCA 7, SCA 8, or unidentified disease-causing mutations. J Neurol 2001;248:870-876.