Hum Genet (2000) 106 : 597–604
Digital Object Identifier (DOI) 10.1007/s004390000320

ORIGINAL INVESTIGATION

**Priyadarshi Basu · Biswanath Chattopadhyay · Prasanta K. Gangopadhaya · Subhas C. Mukherjee
Krishna K. Sinha · Shyamal K. Das · Susanta Roychoudhury · Partha P. Majumder · Nitai P. Bhattacharyya**

Analysis of CAG repeats in SCA1, SCA2, SCA3, SCA6, SCA7 and DRPLA loci in spinocerebellar ataxia patients and distribution of CAG repeats at the SCA1, SCA2 and SCA6 loci in nine ethnic populations of eastern India

Received: 11 January 2000 / Accepted: 25 April 2000 / Published online: 1 June 2000

Abstract To identify various subtypes of spinocerebellar ataxias (SCAs) among 57 unrelated individuals clinically diagnosed as ataxia patients we analysed the SCA1, SCA2, SCA3, SCA6, SCA7 and DRPLA loci for expansion of CAG repeats. We detected CAG repeat expansion in 6 patients (10.5%) at the SCA1 locus. Ten of the 57 patients (17.5%) had CAG repeat expansion at the SCA2 locus, while four had CAG expansion at the SCA3/MJD locus (7%). At the SCA6 locus there was a single patient (1.8%) with 21 CAG repeats. We have not detected any patient with expansion in the SCA7 and DRPLA loci. To test whether the frequencies of the large normal alleles in SCA1, SCA2 and SCA6 loci can reflect some light on prevalence of the subtypes of SCAs we studied the CAG repeat variation in these loci in nine ethnic sub-populations of eastern India from which the patients originated. We report here that the frequency of large normal alleles (>31 CAG repeats) in SCA1 locus to be 0.211 of 394 chromosomes studied. We also report that the frequency of large normal alleles (>22 CAG repeats) at the SCA2 locus is 0.038 while at the SCA6 locus frequency of large normal alleles (>13 repeats) is 0.032. We discussed our data in light of the distribution of normal alleles and prevalence of SCAs in the Japanese and white populations.

P. Basu · B. Chattopadhyay · N. P. Bhattacharyya (✉)
Crystallography and Molecular Biology Division,
Saha Institute of Nuclear Physics,
1/AF, Bidhan Nagar, Calcutta 700 064, India
e-mail: nitai@hp2.saha.ernet.in,
Tel.: +91-33-3370379 ext 311, Fax: +91-33-3374637

P. K. Gangopadhaya · S. K. Das
Bangur Institute of Neurology,
52/1A, Sambhunath Pandit Street, Calcutta, India

S. C. Mukherjee
Medical College Hospital, Calcutta, India

K. K. Sinha
Advanced Diagnostic Centre, Ranchi, India

S. Roychoudhury
Indian Institute of Chemical Biology, Calcutta, India

P. P. Majumder
Indian Statistical Institute, Calcutta, India

Introduction

The spinocerebellar ataxias (SCAs) are autosomal dominant late-onset progressive neurodegenerative disorders displaying clinical and genetic heterogeneity. The loci responsible for seven subtypes of SCAs have been cloned: SCA1 (MIM 164400), SCA2 (MIM 183090), Machado-Joseph disease (MJD)/SCA3 (MIM 109150), SCA6 (MIM 183086), SCA7 (MIM 164500), SCA8 (MIM 271245) and DRPLA (MIM 125370). In all cases, expansion of CAG/CTG repeats in the respective genes have been implicated in the pathogenesis of the disease (David et al. 1997; Imbert et al. 1996; Kawaguchi et al. 1994; Orr et al. 1993; Koide et al. 1994; Koob et al. 1999; Pulst et al. 1996; Sanpei et al. 1996; Zhuchenko et al. 1997). For example, the number of CAG repeats at the SCA1 locus varies from 25 to 36 in normal individuals, while among affected individuals the range is 40 to 81 (Orr et al. 1993). Similarly, polymorphic CAG repeats in the ataxin2 gene (SCA2 locus) varies from 15 to 29 repeats among normal individuals and from 35 to 59 among affected individuals (Imbert et al. 1996; Sanpei et al. 1996).

Wide global variation in relative prevalence of SCA subtypes among autosomal dominant cerebellar ataxia (ADCA) patients has been observed. SCA1 has been reported to be far more common in Russia (Illarioshkin et al. 1996) than any other SCA subtypes. Recently Takano et al. (1998) have reported that the general prevalence of SCA1 and SCA2 is significantly higher among white SCA pedigrees (15% and 14%, respectively) than in the Japanese, whereas relative prevalence of SCA3 is higher in the Japanese pedigrees (43%) than in whites (30%). Also, SCA6 and DRPLA appears to be less frequent in white populations (5% and 0%, respectively) than in Japanese populations (11% and 20%, respectively; Takano et al. 1998).

It has been reported that SCA2 is exclusively responsible for all ataxia cases in the Indian population (Wadia et al. 1997). In a study of six Indian SCA2 pedigrees, Wadia et al. observed CAG repeat expansion in 14 affected family members at the SCA2 locus. All of these patients showed slow saccades and peripheral neuropathy. An inverse correlation between repeat size and age at onset was observed with repeat numbers varying from 36 to 45 repeats (Wadia et al. 1998). Similar observation has been reported by Saleem et al. (2000) in an independent set of 39 SCA pedigrees principally from northern India. SCA2 is also the most common form of hereditary ataxias among Korean patients (Jin et al. 1999), constituting 12.6% of all SCA patients, followed by SCA6 (6.9%) and SCA3 (4.6%). SCA1 or SCA7 mutation was not found in this study.

SCA3 is more common in Germany (Schöls et al. 1995), Brazil (Lopes-Cendes et al. 1997), the United States (Geschwind et al. 1997; Moseley et al. 1998), Portugal (Silveira et al. 1998) and Japan (Sasaki and Tashiro 1999; Takano et al. 1998). In Portugal, expansion at SCA3/MJD locus was observed in 74% of ADCA patients, followed by expansion at SCA2 locus in 4% of patients. The investigators did not find any SCA1, DRPLA or SCA6 mutation (Silveira et al. 1998). Similarly, in the Japanese population, Sasaki et al. (1999) have observed that 24.6% of patients possessed CAG expansion at the SCA3 locus, followed by expansions at SCA6 (11.8%), SCA1 (10.5%) and SCA2 (4.4%) loci. SCA6 is the most common (5%) expansion mutation in SCA patients in the United Kingdom, followed by SCA2 (Leggo et al. 1997). Expansion at the SCA6 locus is also very frequent (13%) in ADCA families of Germany (Schöls et al. 1998).

The molecular basis of such differences in the prevalence of SCAs within and among ethnic populations is unclear. Recent studies on myotonic dystrophy (DM), another trinucleotide repeat expansion disorder, have shown that the distribution of large CTG repeat alleles in the normal range in populations is positively correlated with disease prevalence. This is a characteristic feature of dynamic mutations in general, where the size of an initial expansion determines the rate of further expansion. This has led to the postulation that alleles with large CTG repeat number in the normal range (large ANs) are relatively unstable and undergo expansion to reach an intermediate range from which further expansion to the disease range takes place. Therefore a study on the percentage of large normal alleles in any population would be an indirect reflection of the prevalence of the disease in that population. This was found to be true in African and Israeli populations (Goldman et al. 1996; Mor-Cohen et al. 1997). Similarly, in Huntington's disease it has been postulated that the expanded CAG repeats in HD patients arise from large normal alleles of more than 20 repeats (Rubinsztein et al. 1994; Squiteiri et al. 1994). For example, while the prevalence of Huntington's disease is 1 person per 10,000 in East Anglia (United Kingdom), in the Japanese population it is relatively rare (1 in 1,000,000). This observation is consistent with the fact that normal Japanese individuals show significant paucity of large alleles at the Huntington's disease locus compared to East Anglians. A similar study involving white and Japanese ADCA patients and normal individuals observed that the prevalence of SCAs is highly correlated with the frequency of larger normal alleles in the normal population (Takano et. al. 1998). They reported that prevalence of SCA1 and SCA2 among SCA patients are significantly higher in white SCA pedigrees (15% and 14% respectively) than in Japanese, corresponding to the observation that the frequencies of large ANs in SCA1 (>30 repeats and >31 repeats) and SCA2 (>22 repeats) were significantly higher in whites than in Japanese. This was also consistent with the SCA3, SCA6 and DRPLA loci (Takano et al. 1998).

To our knowledge, there are no previous reports of SCA6 from India, clinical or otherwise, although there are reports of SCA2 and MJD/SCA3 (Bharucha et al. 1986; Wadia and Swami 1971).

In this study we present the results of DNA analysis of CAG repeats at the SCA1, SCA2, SCA3/MJD, SCA6, SCA7 and DRPLA loci in 57 unrelated patients with SCA or late-onset cerebellar ataxia for determining the subtype of SCA in these individuals. We also provide the results of studies on CAG repeat distributions at the SCA1, SCA2 and SCA6 loci in nine ethnic populations of eastern India from which the patients originated, in an attempt to understand the molecular determinants of the ethnic variation in subtypes of SCAs among these populations.

Materials and methods

Patient sampling

Among patients visiting the Bangur Institute of Neurology, Medical College Hospital, Calcutta, and Advanced Diagnostic Centre, Ranchi, during the period from August 1997 to September 1999, 57 were clinically examined and identified as having SCA. These patients were unrelated to one another at least to the second-cousin level. Information of family history and a sample of blood was collected, with appropriate consent, from each patient. Most patients belonged to the Hindu caste groups of eastern India (West Bengal and Bihar). These caste groups were at various levels in the social hierarchy. Nine patients were from Islamic religious groups, but originated from eastern India. None of the patients belonged to any tribal community. Of the 57 patients 26 reported positive family history. The family history of the remaining 31 cases was not sufficient to allow classification as sporadic or late-onset cerebellar ataxia with positive family history. Of the 26 patients with positive family history, transmission was paternal in 15 and maternal in 8, while in 3 others there was no parental history although other family members with similar symptoms were reported. The age at presentation of the disease varied from 9 to 62 years. The mean age at presentation of the patients with positive family history was 34.42±13.04 and that of the remaining patients was 36.83±13.34. There were 37 men and 20 women.

DNA isolation and PCR amplification

DNA was isolated from peripheral blood lymphocytes using standard protocol (Miller et al. 1988). SCA1 locus was amplified by PCR using primers Rep1 and Rep2 under the conditions described by Orr et al. (1993). SCA2, SCA3, SCA6, SCA7 and DRPLA loci

were amplified using primers F1 and R1, MJD52 and MJD25, S-5-F1 and S-5-R1, 4U1024 and 4U716, B37F and B37R, respectively. Conditions for amplification were essentially same as described earlier (Kawaguchi et al. 1994; Koide et al. 1994; Sanpei et al. 1996; Stevanin et al. 1998; Zhuchenko et al. 1997). In the case of patients, the amplified products of all loci were checked on 1.5% agarose gel and then run on 6% denaturing urea-polyacrylamide gel along with positive controls and allelic markers of known sizes. The products were then transferred to Hybond N+ membrane (Amersham, UK), hybridised with $(CTG)_{10}$ oligonucleotide end-labelled with ^{32}P (BRIT, India) by terminal transferase (GIBCO BRL, USA) and visualised by autoradiography. PCR products of the SCA2, SCA3 and SCA7 loci were also run on 2% agarose gel (high-efficiency grade; USB, Amersham, UK) as well and transferred to Hybond N+ membrane and probed with ^{32}P-labelled $(CTG)_{10}$ as described above, to detect any expanded allele which could not be resolved on 6% polyacrylamide gel. For determination of repeat numbers at the SCA2 and SCA6 loci, we ran the PCR products of various normal individuals on 7% non-denaturing polyacrylamide gel with 25-and 50-bp ladders (GIBCO BRL, USA) and visualised by silver staining. One of the normal allele at the SCA2 locus was sequenced and found to have 22 CAG repeats. For determination of sizes of PCR products at the SCA1, SCA3, SCA7 and DRPLA loci, PCR amplification was carried out using fluorescence-labelled primers and detected using Genescan system (version 2.02) in an ABI-377 automated DNA sequencer. While determining the of sizes of normal alleles, we included the cryptic CAT or CAA interruptions which may be present within the CAG tract in SCA1 or SCA2 loci (Orr et al. 1993; Sanpei et al. 1996). One of the expanded alleles at the SCA1 locus was sequenced and had 49 CAG repeats with no CAT interruptions and was used as a positive control. One of the expanded alleles at the SCA3 locus was also sequenced and had 79 CAG repeats.

Population sampling

A sample of 5–10 ml blood was collected by venipuncture from each individual belonging to nine endogamous ethnic groups of India with appropriate consent. DNA was isolated and analysed as described above. The populations inhabit the states of West Bengal, Orissa and Tripura. The ethnic groups are: Bengali Brahmin, Hindu caste, traditionally priests, now in various occupations; Santal, tribal, primarily agricultural labourers; Lodha, tribal, hunter-gatherers and agricultural labourers; Tripuri, tribal, agriculturists; Gaud, Hindu caste, primarily agriculturists; Tanti, Hindu caste, traditionally weavers; Agharia, Hindu caste, primarily agriculturists; Mahishya, Hindu caste, primarily agricultural labourers; Bagdi, Hindu caste, agricultural labourers and fishermen. Of these groups, the Lodhas and Santals are Austro-Asiatic speaking tribals, and the Tripuris are a Tibeto-Burman speaking tribal group. The remaining six populations are Indo-European speaking caste groups at various levels of social hierarchy. Further details about the populations are described in Majumder et al. (1999).

Degrees of freedom was 1 for all χ^2 tests performed between any pair of populations. The null hypothesis was rejected at $P<0.05$.

Results

CAG repeats in SCA1, SCA2, SCA3, SCA6, SCA7 and DRPLA loci in SCA patients

We studied CAG repeat length at the SCA1, SCA2, SCA3, SCA6, SCA7 and DRPLA loci in 57 unrelated ataxia patients for determining the subtypes of SCA. Only 6 patients had CAG repeat numbers in the range generally observed among SCA1 patients. The size of expanded alleles varied from 44 to 52 repeats, with age at presentation between 20 and 53 years. Transmission appears to have been paternal for five patients, and there was no definite history of SCA among relatives of the sixth patient (with 25 and 49 CAG repeats). The expanded allele in this patient was sequenced, and no CAT interruption was detected.

At the SCA2 locus we detected moderate expansion of CAG repeats in 10 of the 57 patients studied (39–45 repeats), with age at presentation varying from 20 to 50 years. There was no definite family history in three of the patients showing SCA2 expansion, whereas transmission was paternal in four cases and maternal in three. Blood samples from the affected parents were available in three cases, and expanded CAG repeat numbers varied from 36 to 40 repeats in these cases.

The region containing the CAG repeats in the ataxin3 gene was amplified and analysed as described above. In 4 of the 57 ataxia patients we observed CAG repeats in the expanded range. One of the expanded alleles was sequenced and had 79 CAG repeats. Transmission was maternal in three cases, with age of presentation varying from 20 to 29 years, while in the fourth case, in which the age at presentation was 62 years, there was no family history of the disease.

At the SCA6 locus we detected only one individual with 21 CAG repeats, which is in the expanded range (20–29 repeats). The homologous chromosome had 11 CAG repeats. Transmission was paternal as revealed from the interview of the patient. In the remaining 56 patients the CAG repeat number varied from 7 to 14 at this locus.

We detected no expansion in the DRPLA and SCA7 loci among these 57 patients. A summary of the results is presented in Table 1.

Table 1 Summary of the distribution of expanded CAG repeats in various SCA loci among 57 clinically diagnosed unrelated ataxia patients

Locus	Expansion of CAG/CTG repeats observed in patients		Total		Range of expanded repeat numbers
	With family history	Without family history	*n*	%	
SCA1	5	1	6	10.5	44–52
SCA2	7	3	10	17.5	39–45
SCA3/MJD	3	1	4	7.0	62–79
SCA6	1	0	1	1.8	21
Unknown	10	26	36	63.2	–

[a]No expansion was obtained at SCA7 and DRPLA loci among the patients studied

CAG repeats at SCA1, SCA2 and SCA6 loci in ethnic populations

The repeat sizes of the normal population samples were determined at the SCA1 locus and is presented in Table 2. The overall estimated heterozygosity considering all the nine populations (394 chromosomes) was 0.84 at the SCA1 locus, with allele range varying from 23 to 36 repeats. There were 14 alleles, with the 29-repeat allele having the maximum frequency. The distribution of alleles is positively skewed (0.26) at this locus. Between populations the observed heterozygosity varied from 0.72 in the Agharias to 0.85 in the Brahmins and Tripuris. Alleles with the maximum frequency varied from 29 to 31 repeats in all populations, except Lodhas, among whom the 34-repeat allele was the most common. The maximum number of alleles (11) was found in the Brahmins and the Tripuris, while the Gaud and Agharias had the minimum number (5). The frequency of alleles greater than 30 or 31 repeats (the large normal alleles or ANs) varied among populations. The frequency of large alleles higher than 31 was 0.21 in the overall sample studied. The χ^2 test was performed between all pairs of populations to determine whether the variation in frequencies of large ANs (>31 repeats) was significant between any pair. Differences between Lodhas and other populations, except the Brahmins, were all statistically significant ($P<0.05$). The frequency was also significantly higher in Brahmins than in Agharias ($\chi^2=5.96$; $P<0.015$).

The observed distributions of repeat sizes at the SCA2 locus are presented in Table 3. In the normal population the observed heterozygosity in the pooled sample of 448 chromosomes was 0.17, with allele range varying from 14 to 31 repeats. The distribution of alleles was negatively skewed (–0.13) at this locus. We obtained one allele each with 31 CAG repeats in the Brahmin and Mahishya populations, which is above the range of the initially published maximum allele size of 29 repeats (Sanpei et al. 1996). There were 12 alleles in all, with the 22-repeat allele having the maximum frequency (0.91). Among the populations the heterozygosity varied from 0.36 in the Lodhas to 0.0 in the Gaud population. The 22-repeat allele had the maximum frequency in all populations. The maximum number of alleles was found among Brahmins (7), while the Gaud population was monoallelic for the 22-repeat allele. The frequency of alleles greater than 22 repeats (the large normal alleles or ANs) vary among populations. The

Table 2 Analysis of distribution of CAG repeat numbers of the SCA1 locus in nine ethnic populations

Population	2*n*	CAG repeats	Observed heterozygosity	CAG repeat range: chromosomes at each class interval (%)			Allele range (in repeats)	Total no. of alleles	Allele with maximum repeats
				23–26	27–31	32–36			
Brahmin	44	30.64±2.40	0.85	6.8	61.5	31.7	23–35	11	30
Santal	38	30.39±2.28	0.80	0.0	84.1	15.9	27–36	8	31
Lodha	62	31.52±2.18	0.77	0.0	56.5	43.5	27–35	8	34
Tripura	72	29.43±2.30	0.85	4.2	76.3	19.5	23–35	11	29
Gaud	14	29.71±1.68	0.76	0.0	92.8	7.2	27–32	5	31
Tanti	22	29.36±1.92	0.74	0.0	91.0	9.0	27–35	6	29
Agharia	26	29.65±1.67	0.72	0.0	96.2	3.8	27–32	5	31
Mahishya	56	30.05±1.89	0.79	3.6	82.1	14.3	25–34	9	29
Bagdi	60	30.20±2.20	0.82	1.7	81.6	16.7	26–36	10	29, 31
Pooled	394	30.21±2.23	0.84	2.2	76.7	21.1	23–36	14	29

Table 3 Analysis of CAG repeats in the SCA2 locus in nine ethnic populations of India

Population	2*n*	CAG repeats	Observed heterozygosity	CAG repeat range: chromosomes at each class interval (%)			Allele range (in repeats)	Total no. of alleles
				<22	22	>22		
Brahmin	44	22.27±1.75	0.29	4.5	84.1	11.4	17–31	7
Santal	40	22.12±0.78	0.05	0.0	97.5	2.5	22–27	2
Lodha	64	21.53±0.92	0.36	21.9	78.1	0.0	19–22	4
Tripuri	80	21.89±1.44	0.14	3.7	92.6	3.7	14–27	6
Gaud	24	22 ±0.0	0.0	0.0	100.0	0.0	22	1
Tanti	30	22.07±0.24	0.13	0.0	93.3	6.7	22–23	2
Agharia	36	22.06±0.33	0.05	0.0	97.2	2.8	22–24	2
Mahishya	68	22.19±1.14	0.09	0.0	95.5	4.5	22–31	4
Bagdi	62	21.74±1.24	0.18	6.5	90.3	3.2	16–23	5
Total	448	21.95±1.15	0.17	5.2	91.0	3.8	14–31	12

Table 4 Analysis of CAG repeats in the SCA6 locus in nine ethnic populations of India

Population	2*n*	CAG repeats	Observed heterozygosity	CAG repeat range: chromosomes at each class interval (%)				Allele range (in repeats)	Total no. of alleles	Allele with maximum frequency (in repeats)
				7	8–10	11–13	>13			
Brahmin	44	10.68±2.140	0.71	22.7	0.0	77.3	0.0	7–13	4	11
Santal	40	10.65±1.838	0.52	17.5	0.0	82.5	0.0	7–13	3	11
Lodha	64	11.86±1.309	0.65	3.1	0.0	95.3	1.6	7–15	5	11
Tripuri	78	11.21±2.488	0.76	21.8	2.6	66.6	9.0	7–15	7	13
Gaud	16	11.62±1.568	0.76	12.5	0.0	75.1	12.4	7–16	6	11
Tanti	26	10.58±1.843	0.63	19.2	0.0	80.8	0.0	7–13	4	11
Agharia	32	11.60±1.453	0.70	6.2	0.0	93.8	0.0	7–13	4	11
Mahishya	56	11.30±1.812	0.69	10.7	1.8	82.1	5.4	7–15	7	11
Bagdi	58	10.88±1.713	0.60	13.8	0.0	86.2	0.0	7–13	4	11
Total	414	11.17±1.962	0.71	14.2	0.7	81.9	3.2	7–16	9	11

maximum frequency in that range was found among the Brahmins (0.114) while the minimum was found among the Lodhas and Gauds (0.0). The frequency is 0.038 in the overall population out of 448 chromosomes studied. The χ^2 test performed between all pairs of populations revealed that the differences were significant between Brahmins and Lodhas (χ^2=7.25, P<0.01) and between Lodhas and Tantis (χ^2=4.25, P<0.05). The Brahmin was the only population in which the frequency of large ANs of more than 22 repeats was significantly higher than the pooled population (χ^2=5.12; P<0.05). Considering the mean of CAG repeat alleles in various populations, the standard deviation from the mean (22.27) was also the highest in the Brahmins (Table 3). The mean of CAG repeats was 21.95 in the pooled population, with a standard deviation of 1.15.

The repeat numbers at the SCA6 locus were obtained in the same set of populations, and the results are presented in Table 4. Analysis of 414 chromosomes in the pooled sample revealed that the observed heterozygosity was 0.71, with the 11-repeat allele having the maximum frequency. The distribution of alleles was positively skewed (0.28). In all, nine alleles were found ranging from 7–16 CAG repeats. There was no 9-repeat allele in any population although 8- and 10-repeat alleles were found. Among the nine populations studied the heterozygosity varied from a maximum of 0.76 among Gauds and Tripuris to a minimum of 0.52 among Santals. The 11-repeat allele had the maximum frequency in all populations, except among Tripuris, in whom the 13-repeat allele was the most frequent. The maximum number of alleles were found in the Tripuris and Mahishyas (7), while the Santals had the minimum number of alleles (3). Alleles ranging from 8 to 10 repeats were very rare and were present only in the Tripuris and Mahishyas. The frequency of alleles greater than 13 repeats (the large ANs) varied among populations. The maximum frequency for that range was found in the Gaud (0.124), while in the Brahmin, Santal, Tanti, Agharia and Bagdi populations, no alleles in this range was found. The frequency was 0.032 in the overall population of 414 chromosomes studied. The difference in frequencies of large normal alleles (greater than 13 repeats) was significant between Brahmin and Tripuri (χ^2= 3.94, P<0.05) and between Tripuri and Bagdi (χ^2=5.22, P<0.05). Differences between Gaud and some other populations, namely the Brahmin, Santal, Lodha, Agharia and Bagdi, were all statistically significant (P<0.05).

At the SCA3 locus we studied three populations: the Brahmin, Lodha and Santal. In both Brahmins and Lodhas (n=36 and 48, respectively) the range of CAG repeats varied from 14 to 30 repeats, with 11 and 12 alleles in the two populations, respectively. The Santals (n=28) had 14–32 repeats with 10 distinct alleles. The 23-repeat allele was the most frequent in all three populations. Heterozygosity at this locus varied from 0.82 in the Brahmin and Lodha to 0.87 in the Santal. The overall frequency of large ANs with more than 27 and more than 31 repeats was 0.17 and 0.026, respectively, in the pooled population (112 chromosomes).

Discussion

The relative frequencies of the hereditary ataxias vary worldwide and also among different ethnic groups. Contrary to an earlier observation that SCA2 is exclusively responsible for all ataxia cases in the Indian population (Wadia et al. 1997), our study clearly indicates that at least in populations from eastern India various other subtypes of SCAs are also present. The frequencies of SCAs as detected by the CAG repeat expansion in the respective genes were 10.5%, 17.5%, 7% and 1.8% for SCA1, SCA2, SCA3 and SCA6, respectively. No SCA7 or DRPLA patients were identified by this screening. To our knowledge, this is the first report on simultaneous molecular analysis of SCA1, SCA3, SCA6, SCA7 and DRPLA among SCA patients and normal individuals from defined ethnic groups of eastern India.

We studied the CAG repeat distribution at the SCA1, SCA2 and SCA6 loci in nine ethnic populations of eastern India from which the patients originated, to examine whether the data suggest predisposition to a particular type

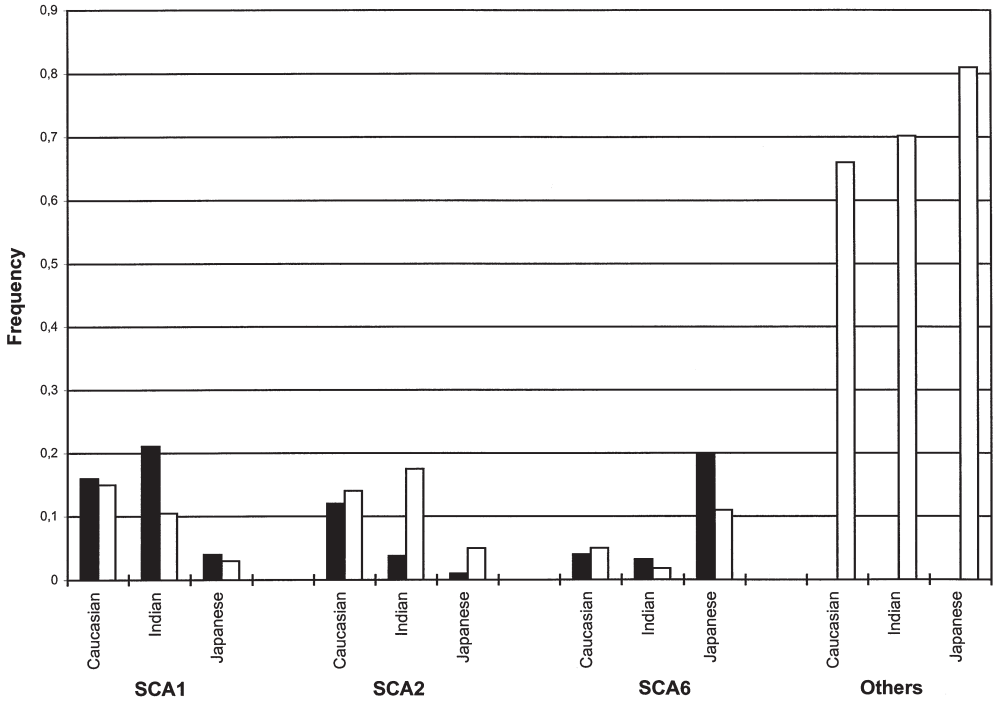

Fig. 1 Histogram representing prevalence and large normal alleles (large ANs) at various SCA loci among different populations. *Black bar* Frequency of large ANs; *white bar* prevalence at various SCA loci in different populations. The large ANs are alleles of more than 31 repeats at the SCA1 locus, alleles of more than 22 repeats at the SCA2 locus and alleles of more than 13 repeats at the SCA6 locus. (Data on white and Japanese populations from Takano et al. 1998)

of SCA. This would also help us to determine whether the prevalence varies between different ethnic groups. Unfortunately, our data on SCA3 among normal individuals is still scanty for this purpose.

At the SCA1 locus, large normal alleles (the large ANs) greater than 30 and 31 repeats constituted 43.1% and 21.1% of 394 chromosomes in the pooled population. In a study by Takano et al. (1998) the frequencies of alleles greater than 30 and 31 repeats were found to be 0.09 and 0.04 among Japanese, while they were 0.26 and 0.16 in whites, respectively. This is well correlated with the observation that prevalence of SCA1 among SCA patients is much higher in white pedigrees (15%) than in Japanese pedigrees (3%). Considering large ANs of more than 30 repeats, the frequency in the pooled Indian population is significantly higher than that of both the Japanese and white populations ($\chi 2=79.62$, $P<0.0001$ between Indian and Japanese; $\chi 2=28.51$, $P<0.0001$ between Indian and white). However, when the frequencies of large ANs of more 31 repeats are compared, the frequency among Indians is significantly higher than among Japanese ($\chi 2=41.75$, $P<0.0001$), whereas there was no significant difference with the white population. This indicates that, as with whites, SCA1 may represent a larger proportion among SCA subtypes in some Indian sub-populations. The frequency of large ANs and the prevalence of the corresponding SCAs in the overall population are shown schematically in Fig. 1.

At the SCA2 locus, if we consider alleles of more than 22 repeat to constitute the large ANs from which further expansion to the disease range takes place, the frequency of large ANs is 0.038 in the pooled population. The maximum frequency of large ANs was found in the Brahmins (0.114) while the minimum was found in the Lodhas and Gaud (0.0). We further compared our data on large ANs with those on Japanese and white populations presented in the study by Takano et al. (1998). The frequencies of alleles of more than 22 repeats in Indian, Japanese and white populations are 0.038, 0.01, 0.12, respectively. The pairwise χ^2 test revealed that the frequency of large ANs is significantly higher in the Indian population (pooled) than in the Japanese population ($\chi^2=10.66$, $P<0.0015$). Similar comparison revealed that the frequency of large ANs is significantly lower in the Indian population than in whites ($\chi^2=20.81$, $P<0.0001$). The frequency of large ANs of more than 22 repeats in our Indian population is in the intermediate range between the two other populations. These observation is consistent with the finding that Indians are genetically between Caucasoids and Mongoloids (Majumder 1998). However, the prevalence of SCA2 in our population appears to be highest among the SCA subtypes (Fig. 1) and is even higher than that among whites. This could be due either to the small patient sample size or to the fact that the total number of large ANs in our pooled population is underrepresented, since at least two populations, the Lodha and the Gaud, do not contribute any large ANs to the pooled data.

We found two 14-repeat alleles at the SCA2 locus in the Tripuri population. This is the smallest allele in our population. The largest allele is of 31 CAG repeats found one each in the Brahmin and Mahishya populations, which is larger than the initially reported 29 repeats (Sanpei et al. 1996).

At the SCA6 locus the frequency of large ANs of more than 13 repeats is 0.032 for the overall population. The maximum frequency of alleles in this range is found among the Gaud (12.4%), while the Brahmin, Santal, Tanti, Agharia and Bagdi do not have any alleles in this

range. As at the SCA1 and SCA2 loci, we compared our data on large ANs with those from Japanese and white populations. The frequencies of alleles of more than 13 repeats in the Indian, Japanese and Caucasian populations are 0.032, 0.2, and 0.04, respectively. The frequency of large ANs (>13 repeats) was significantly lower in the Indian population than in Japanese population (χ^2=53.59, P<0.0001). No significant difference was observed between the Indian and white populations. When alleles of more than 14 repeats were considered, the frequency in the Indian population was significantly lower than that among Japanese (χ^2=23.6, P<0.0001) but higher than that among whites (χ^2=5.75, P<0.05). However, compared to an incidence of SCA6 of 11% and 5%, respectively, among Japanese and white SCA patients, our data show a lower prevalence of SCA6 among SCA patients. This observation is consistent with the fact that most of the ethnic populations studied (Brahmin, Santal, Tanti, Agharia and Bagdi) do not have any large ANs greater than 13 repeats at this locus. It is interesting to note that the Tripuris, who are a Tibeto-Burman speaking tribal group, and the only Mongoloid population studied (they are a part of the Bodo ethnic stock who migrated from Tibet several centuries ago), stands out from the rest at this locus (Table 4). They have the highest heterozygosity and the maximum number of alleles among the study populations. They also have the second highest frequency of large ANs (second to Gaud, Hindu caste group) and, unlike the other populations, have the 13-repeat allele as the most frequent allele. The Tripuris are also the only population in whom the frequency of large ANs of more than 13 repeats is significantly higher than that in the pooled population (χ^2=5.5; P<0.05).

It is also interesting to note that one of the two Austro-Asiatic speaking tribal populations, the Lodhas (a numerically small and geographically isolated group), stands out from the rest considering both the SCA1 and SCA2 loci. They have the highest frequency of large ANs and highest mean CAG length at the SCA1 locus and have the 34-repeat as the most frequent allele, unlike the other populations. At the SCA2 locus they have the highest heterozygosity although they do not have any alleles at the higher range (>22 repeats) and the lowest mean CAG length. It has recently been proposed that these Austro-Asiatic speaking tribal groups represent the ancestral populations who moved to India during one of the early waves of „out-of-Africa“ migration (Majumder et al. 1999). The other tribal group, the Santals, exhibit least heterozygosity and has the least number of alleles at the SCA6 locus.

Our studies indicate that among SCA subtypes, SCA1 and SCA2 are the most frequent in the eastern Indian populations, while SCA6 is less frequent. Our studies also support the notion that larger ANs could be an indirect source of information on population-specific SCAs. This information would be particularly important for populations in which no epidemiological data on prevalence of the disease is available. Further studies on SCA loci, including larger patient samples, are necessary to substantiate this viewpoint, particularly because after classification of the SCA patients included in the present study to the various ethnic groups the sample sizes were too small to establish meaningful correlations between disease prevalence and the frequency of large ANs.

Acknowledgements We are grateful to Dr. Bidyut Roy, Mrs. Monami Roy, Mr. Badal Dey and Mr. Madan Chakraborty for collection of population samples and isolation of genomic DNA. We are thankful to Mr. Utpal Basu for his technical support. The work was partially supported by a grant from the Department of Biotechnology, Government of India, to N.P.B.

References

Bharucha NE, Bharucha EP, Bhaba SK (1986) Machado-Joseph-Azorean disease in India. Arch Neurol 43:142–144

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

Geschwind DH, Perlman S, Figueroa CP, Treiman LJ, Pulst SM (1997) 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 60: 842–850

Goldmann A, Ramsay, Jenkin T (1996) Ethnicity and myotonic dystrophy: a possible explanation for its absence in sub-Saharan Africa. Am J Hum Genet 60:57–65

Illarioshkin SN, Slominsky PA, Ovchinnikov IV, Markova ED, Miklina NI, Klyushnikov SA, Shadrina M, et al (1996) Spinocerebellar ataxia type 1 in Russia. J Neurol 243:506–510

Imbert G, Frederic S, Yvert G, Devys D, Trottier Y, Garnier J-M, et al (1996) Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. Nat Genet 14:285–291

Jin DK, Oh MR, Song SM, Koh SW, Lee M, Kim GM, et al (1999) Frequency of spinocerebellar ataxia types 1:2:3:6:7 and dentatorubral pallidoluysian atrophy mutations in Korean patients with spinocerebellar ataxia. J Neurol 246:207–210

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

Koide R, Ikeuchi T, Onodera O, Tanaka H, Igarashi S, Endo K, et al (1994) Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). Nat Genet 6:14 –18

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

Leggo J, Dalton A, Morrison PJ, Dodge A, Connarty M, Kotze MJ, Rubinztein DC (1997) Analysis of spinocerebellar ataxia types 1:2:3, and 6, dentatorubral-pallidoluysian atrophy, and Friedreich's ataxia genes in spinocerebellar ataxia patients in the UK. J Med Genet 34:982–985

Lopes-Cendes I, Teive HG, Calcagnotto ME, Da Costa JC, Cardoso F, Viana E, et al (1997) Frequency of the different mutations causing spinocerebellar ataxia (SCA1, SCA2, MJD/SCA3 and DRPLA) in a large group of Brazilian patients. Arq Neuropsiquiatr 55:519–529

Majumder PP (1998) People of India: biological diversity and affinities. Evol Anthropol 6:100–110

Majumder PP, Roy B, Banerjee S, Chakraborty M, Dey B, Mukherjee N, et al (1999) Human-specific insertion/deletion polymorphisms in Indian populations and their possible evolutionary implications. Eur J Hum Genet 7:435–446

Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16:1215

Mor-Cohen R, Magal N, Gadoth N, Shohat T, Shohat M (1997) Correlation between the incidence of myotonic dystrophy in different groups in Israel and the number of CTG trinucleotide repeats in the myotonin gene. Am J Med Genet 71:156–159

Moseley ML, Benzow KA, Schut LJ, Bird TD, Gomez CM, Barkhaus PE, et al (1998) Incidence of dominant spinocerebellar and Friedreich triplet repeats among 361 ataxia families. Neurology 51:1666–1671

Orr HT, Chung M, Banfi S, Kwiatkowski TJ, Servadio A, Beaudet AL, et al (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type1. Nat Genet 4:221–226

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

Rubinsztein DC, Amos W, Leggo J, Goodburn S, Ramesar RS, Old J, et al (1994) Mutational bias provides a model for the evolution of Huntington's disease and predicts a general increase in disease prevalence. Nat Genet 7:525–530

Saleem Q, Choudhury S, Mukherji M, Bashyam L, Padma MV, Chakravarthy A, et al (2000) Molecular analysis of autosomal dominant hereditary ataxias in the Indian population: high frequency of SCA2 and evidence for a common founder mutation. Hum Genet 106:179–187

Sanpei K, Takano H, Sato T, Oyake M, Sasaki H, Wakisaka A, et al (1996) Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. Nat Genet 14:227–284

Sasaki H, Tashiro K (1999) Frequencies of triplet repeat disorders in dominantly, inherited spinocerebellar ataxia (SCA) in the Japanese. Nippon Rinsho 57:21–25

Schöls L, Vieira-Saecker AM, Schöls S, Przuntek H, Eppen JT, Reiss O (1995) Trinucleotide expansion within the MJD1 gene presents clinically as spinocerebellar ataxia and occurs most frequently in German SCA patients. Hum Mol Genet 4:1001–1005

Schöls L, Kruger R, Amoiridis G, Przuntek H, Epplen JT, Riess O (1998) Spinocerebellar ataxia 6: genotype and phenotype in German kindreds. J Neurol Neurosurg Psychiatry 64:67–73

Silveira I, Coutinho P, Maciel P, Gaspar C, Hayes S, Dias A, et al (1998) Analysis of SCA1, DRPLA, MJD, SCA2 and SCA6 CAG repeats in 48 Poruguese ataxia famalies. Am J Med Genet 28:134 –138

Squiteiri F, Andrew SE, Goldberg YP, Kremer B, Spence N, Zeisler J, et al (1994) DNA haplotype analysis of Huntington disease reveals clues to the origins and mechanisms of CAG expansion and reasons for geographical variations of prevalence. Hum Mol Genet 3:2103–2114

Stevanin G, Guinti P, Davia G, Balal S, Durr A, Ruberg M, et al (1998) De novo expansion of intermediate alleles in spinocerebellar ataxia 7. Hum Mol Genet 7:1809–1813

Takano H, Cancel G, Ikeuchi T, Lorenzetti D, Mawad R, Stevanin G, et al (1998) Close associations between prevalences 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 63: 1060–1066

Wadia NH, Swami RK (1971) A new form of heredo-familial spinocerebellar degeneration with slow eye movements. Brain 94:359–379

Wadia NH, Chamberlain S, Desai JD, Mankodi AK, Desai MJ, Twells R (1997) Hereditary cerebellar ataxia with slow saccades (SCA2) more prevalent in India. Abstract. Presented at the First International Symposium on Inherited Ataxia, 29 May–1 June, Montreal

Wadia N, Pang J, Desai J, Mankodi A, Desai M, Chamberlain S (1998) A clinicogenetic analysis of six Indian spinocerebellar ataxia (SCA2) pedigrees. The significance of slow saccades in diagnosis. Brain 121:2341–2355

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