ORIGINAL ARTICLE

Maria do Carmo Costa · Paula Magalhães
Laura Guimarães · Patrícia Maciel · Jorge Sequeiros
Alda Sousa

The CAG repeat at the Huntington disease gene in the Portuguese population: insights into its dynamics and to the origin of the mutation

Abstract Huntington disease (HD) is caused by an expansion of a CAG repeat. This repeat is a dynamic mutation that tends to undergo intergenerational instability. We report the analysis of the CAG repeat in a large population sample (2,000 chromosomes) covering all regions of Portugal, and a haplotype study of $(CAG)_n$ and $(CCG)_n$ repeats in 140 HD Portuguese families. Intermediate class 2 alleles represented 3.0% of the population; and two expanded alleles (36 and 40 repeats, 0.11%) were found. There was no evidence for geographical clustering of the intermediate or expanded alleles. The Portuguese families showed three different HD founder haplotypes associated with 7-, 9- or 10-CCG repeats, suggesting the possibility of different origins for the HD mutation among this population. The haplotype carrying the 7-CCG repeat was the most frequent, both in normal and in expanded alleles. In general, we propose that three mechanisms, occurring at different times, may lead to the evolution from normal CAGs to full expansion: first, a mutation bias towards larger alleles; then, a stepwise process that could explain the CAG distributions observed in the more recent haplotypes; and, finally, a pool of intermediate (class 2) alleles more prone to give rise to expanded HD alleles.

M.C. Costa · P. Maciel
School of Health Sciences, Life and Health Sciences Research Institute (ICVS), University of Minho, Braga, Portugal

P. Magalhães · J. Sequeiros · A. Sousa
Institute for Molecular and Cell Biology,
UnIGENe, Porto, Portugal

L. Guimarães
University Fernando Pessoa, Porto, Portugal

J. Sequeiros · A. Sousa (✉)
Department of Populations Studies, ICBAS, University of Porto,
4099-003 Porto, Portugal
E-mail: absousa@icbas.up.pt
Tel.: +351-22-2062252
Fax: +351-22-2062232

Introduction

Huntington disease (HD) is an autosomal dominant progressive neurodegenerative disorder, usually of adult onset, associated with involuntary choreic movements, cognitive impairment and personality changes, leading to depression and temperamental behaviour (Martin and Gusella 1986). HD is caused by the expansion of a CAG repeat in the coding region of the *HD* gene located on chromosome 4p16.3 (The Huntington Collaborative Research Group 1993). The CAG repeat sizes are: 6–26 CAGs in normal stable alleles (class 1); 27–35 CAGs in non-pathogenic, but expandable alleles, upon paternal transmission (class 2); 36–39 CAGs in expanded alleles with reduced penetrance (class 3); and ≥40 CAGs in fully penetrant alleles (class 4) (The American College of Medical Genetics 1998). The expanded CAG repeat is a dynamic mutation that tends to undergo intergenerational instability, leading to increased length of this repeat in the following generations and, consequently, to anticipation of age at onset of disease (Andrew et al. 1993; Duyao et al. 1993). This phenomenon of anticipation, in cases where the onset of disease is before reproductive age, and occasional infantile/juvenile cases, may give rise to the loss of some *HD* genes from the population; however, there is no tendency towards a decrease in HD prevalence over time, meaning that the frequency of mutant *HD* alleles is being maintained in some way. It has been proposed that all lengths of CAG alleles at the *HD* gene are subjected to a mutational bias towards larger alleles, and that the mutation rate is strongly dependent on the size of the repeat (Rubinsztein et al. 1994). In addition, haplotype analyses in HD families of different ethnic populations has suggested that multiple mutation events underlie this disorder, and

that a specific group of normal unstable alleles within a particular haplotype may be more prone to expansion leading to fully penetrant *HD* alleles, thus constituting a pool for the generation of new *HD* genes (Goldberg et al. 1993; Squitieri et al. 1994; Almqvist et al. 1995). Previously, from our routine genetic testing in Portugal, we observed that normal unstable alleles (class 2) had a relatively high frequency, accounting for 3.7% of the "normal" chromosomes (Costa et al. 2003).

In the present study, we report the analysis of the CAG repeat in a large population sample (2,000 chromosomes) covering all regions of Portugal (including the Azores and Madeira islands), and a haplotype study of $(CAG)_n$ and $(CCG)_n$ repeats in 140 HD Portuguese families. The aims of this study were: (1) to characterise the distribution of the CAG repeat alleles in the Portuguese population; (2) to test previously reported models for the evolution of the HD CAG repeat; (3) to study the origins of the HD mutation in Portuguese families; and (4) to test for a potential effect of the CCG repeat size on age at onset of HD.

Materials and methods

Subjects

The control sample was obtained from a random and anonymous collection of 1,000 Guthrie cards from individuals born in 1998, composed of 50 individuals from each of the 20 regions of Portugal: Aveiro, Azores islands, Beja, Braga, Bragança, Castelo Branco, Coimbra, Évora, Faro, Guarda, Leiria, Lisbon, Madeira island, Portalegre, Porto, Santarém, Setúbal, Viana do Castelo, Vila Real and Viseu. Previous consent was obtained from the Ethics Commission of the Medical Genetics Institute from where samples were obtained. To characterise the distribution of normal alleles from heterozygous HD patients, we used a sample comprising one affected individual, randomly selected from each family. A total of 140 HD Portuguese families, consisting of 238 individuals who were either patients or relatives, were used for haplotype studies.

Methods

Genomic DNA for molecular diagnosis was isolated from peripheral blood using standard protocols (Sambrook et al. 1989). For the control population DNA was isolated from Guthrie cards using Chelex (BioRad) at a final concentration of 0.8%. The CAG and CCG repeat sizes were assessed for all samples by the use of polymerase chain reaction (PCR) analysis with primers that flank either the CAG repeat or the CCG repeat or both (Andrew et al. 1994). The size of the PCR products was determined by denaturing 6% polyacrylamide gel electrophoresis, in parallel with an M13 sequence ladder, and visualised by means of autoradiography. As a positive control for PCR, we used genomic DNA from a patient with an expanded allele containing 94 CAGs in every reaction.

Statistical analysis

Analyses of the CAG repeat allele distribution were performed for the total control population and for each region individually. Descriptive statistics relevant to the CAG repeat (frequencies of allele and classes of allele, range, mean, median, mode, skewness and kurtosis) were determined using the SPSS ver. 12.0 package (SPSS, Chicago, IL). Normality of the CAG repeat distribution was tested by the Kolmogorov–Smirnov test. Chi-square was used to test different proportions of classes of alleles (1, 2, 3 and 4) by region (SPSS 12.0). Student's *t*-test was used to compare mean age at onset between patients with different expanded haplotypes carrying the CCG_7 or the CCG_{10} repeat (SPSS 12.0). Fisher's exact test (2×2) was performed to compare the repeat proportions of CCG_7 and CCG_{10} between normal and mutated HD alleles. Hardy–Weinberg equilibrium analysis was performed using the exact test developed by Guo and Thompson (1992), with the Genepop 3.4 package (Raymond and Rousset 1995). Heterozygosity estimates were based on the observed and the expected number of heterozygotes (using Levene's correction), calculated by the Genepop 3.4 package (Raymond and Rousset 1995). Population differences among the Portuguese regions were assessed with the Fisher's exact test, as implemented in Genepop Version 3.4 (Raymond and Rousset 1995).

Results

Distribution of CAG repeat alleles in the control population

Among the total control population of Portugal (1,000 Guthrie cards), a CAG repeat size in the *HD* gene was obtained for 1,772 chromosomes. The distribution of the $(CAG)_n$ size did not fit a normal distribution ($Z = 9.311$; $P < 0.001$) (Fig. 1). The $(CAG)_n$ tract ranged between 9 and 40 repeats, with the 17 CAG allele occurring most frequently (37.9%). Intermediate alleles of class 2 (27–35 CAGs) represented 3.0% of the control population. Two expanded alleles (0.11%) were found: a repeat allele with reduced penetrance (36 CAGs) present in Guarda, and a fully penetrant repeat (40 CAGs) found in the Azores islands. The exact test showed no significant deviation to the frequencies expected for the Hardy–Weinberg equilibrium ($P = 0.3780 \pm 0.0148$); a heterozygosity of 81.0% was observed (Table 1).

The characterisation of $(CAG)_n$ repeat distribution for each of the 20 regions is shown in Table 1. Analysis of each region individually revealed that the majority

Fig. 1 Distribution of the $(CAG)_n$ size of the Huntington disease (*HD*) gene in 1,772 chromosomes of the general Portuguese population
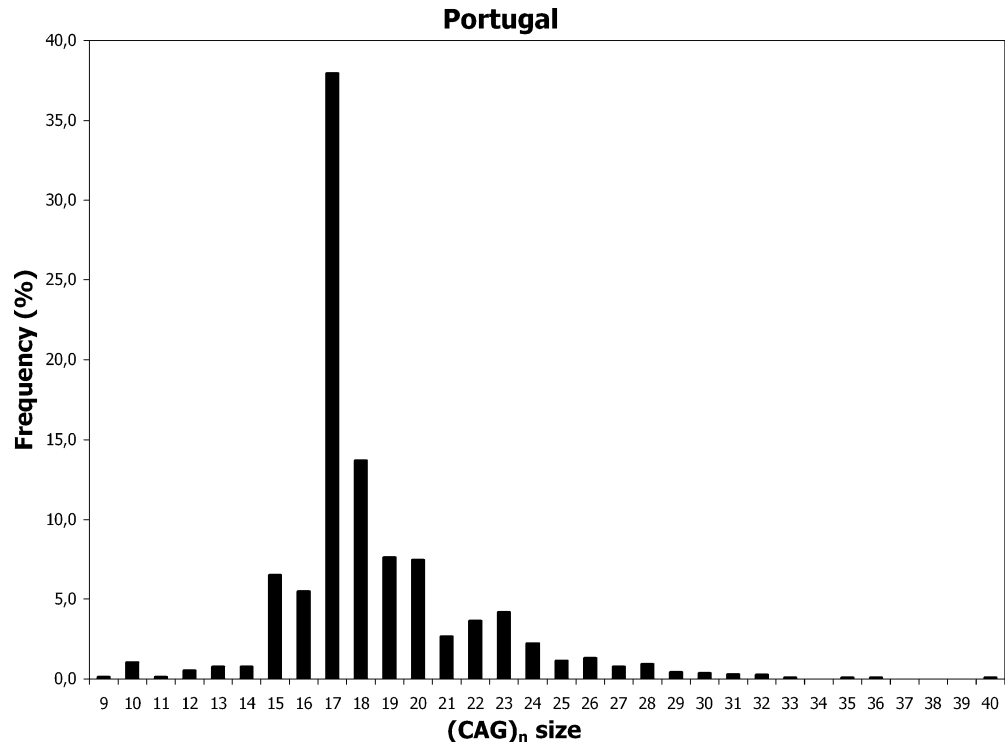
was in Hardy–Weinberg equilibrium, with the exception of Bragança ($P = 0.0223 \pm 0.0039$), and Viana do Castelo ($P = 0.0242 \pm 0.0039$), which showed a significant deviation; however, no significant differences were found between the observed and expected heterozygosity in these two regions. Heterozygosity ranged between 70.6 and 90.0%. There were no differences among all regions, either in relation to $(CAG)_n$ allele distribution ($P = 0.0727 \pm 0.0111$), or to genotype distribution ($P = 0.0989 \pm 0.0143$). In each region, normal unstable (class 2) alleles ranged from 0% in Santarém to 6.3% in Setúbal. The proportion of each class of alleles (classes 1, 2, 3 and 4) was also identical among regions.

Haplotype analysis in HD families

In each of the 140 HD families, one individual carrying the affected chromosome was randomly selected. For these probands, we have constructed haplotypes based on the determination of the size of both the CAG repeat (Fig. 2a) and the CCG repeat. The distribution of the normal CAG repeat alleles from this sample and that from the total control chromosomes did not show significant differences. The distributions of the CAG allele classes between these two populations were also similar. In relation to the CCG repeat, we found four different alleles with size of 7, 8, 9 and 10 repeats. The 7-CCG and the 10-CCG repeat were the most frequent, both in normal (61.2 and 28.8%) and in mutated (87.9 and 11.3%) alleles (Table 2).

Analysis of CCG repeat haplotypes revealed that the HD expanded alleles presented three different haplotypes,

carrying the 7, 9 or 10-CCG allele (Table 2). The 8-CCG was associated only with normal chromosomes (5.8%) (Fig. 2c). Comparison of the proportion of 7-CCG and 10-CCG repeats between normal (< 36 CAGs) and mutant alleles (≥ 36 CAGs) showed an association of the 7-CCG repeat with expanded $(CAG)_n$ alleles, and a more frequent presence of the 10-CCG repeat together with normal CAG repeats ($P < 0.001$). Analyses of the distributions of CAG repeat lengths of chromosomes carrying each of the four CCG repeats showed several differences (Fig. 2): (1) the distribution for the 7-CCG allele proved more representative of the whole sample (Fig. 2b); (2) distributions for the alleles with 10 and, especially, 8 and 9 CCGs, revealed very little variability and no evidence for any association with class 2 alleles (Fig. 2c–e). In normal chromosomes, the 7- and the 10-CCGs alleles were associated with a mean CAG repeat length of 18.68 ± 3.31 and 16.93 ± 1.99 ($P < 0.001$). In expanded chromosomes, the 7 and the 10-CCGs alleles were associated with a mean $(CAG)_n$ size of 46.12 ± 8.37 and 43.90 ± 3.78, respectively (not significantly different; $P = 0.412$).

A mean age at onset of 42.2 years was observed in individuals carrying the 7-CCG haplotype associated with the expanded CAG repeat ($n = 86$), i.e. lower, but not significantly different ($P = 0.057$), from that of patients with the 10-CCG haplotype (mean age at onset = 51.4 years; $n = 10$). Additionally, no effects of the size of the CCG repeat in normal alleles on age at onset of disease were observed, since patients carrying the 7-CCG in the normal chromosome ($n = 60$) showed a mean age at onset of 42.6 years, and those having the 10-CCG ($n = 31$) presented a mean age at onset of 44.5 years.

Table 1 Characterisation of the Huntington disease (*HD*) CAG repeat distribution in Portugal and in each of its regions. *HW* Hardy–Weinberg equilibrium

Population	Alleles (*N*)	(CAG)$_n$ range	(CAG)$_n$ repeat classes (<27) n	(%)	(27–35) n	(%)	(36–39) n	(%)	(>39) n	(%)	Mean ±SE	Median	Mode	Skewness ± SE	Kurtosis ± SE	HW P-value ± SE	Heterozygosity observation/ experiments
Azores islands	100	10–40	96	(96.0)	3	(3.0)	–	–	1	(1.0)	19.17±0.38	18.00	17	2.13±0.24	9.10±0.48	0.4622±0.0137	0.900/0.859
Aveiro	96	10–28	95	(99.0)	1	(1.0)	–	–	–		18.05±0.27	17.00	17	0.84±0.25	2.68±0.49	0.8547±0.0099	0.750/0.746
Beja	86	10–31	83	(96.5)	3	(3.5)	–	–	–		18.59±0.38	17.00	17	0.99±0.26	1.92±0.51	0.4265±0.0138	0.930/0.852
Braga	84	10–28	83	(98.8)	1	(1.2)	–	–	–		18.24±0.32	17.00	17	0.87±0.26	1.78±0.52	0.5718±0.0141	0.738/0.776
Bragança	94	12–30	89	(94.7)	5	(5.3)	–	–	–		18.30±0.35	17.00	17	1.65±0.25	2.80±0.49	0.0223±0.0039*	0.723/0.811
Castelo Branco	82	10–31	81	(98.8)	1	(1.2)	–	–	–		17.99±0.35	17.00	17	0.75±0.27	3.94±0.53	0.1414±0.0085	0.756/0.823
Coimbra	68	9–27	67	(98.5)	1	(1.5)	–	–	–		18.18±0.35	17.00	17	0.78±0.29	2.15±0.57	0.4719±0.0117	0.706/0.761
Évora	90	10–28	86	(95.6)	4	(4.4)	–	–	–		18.74±0.34	18.00	17	0.76±0.25	1.82±0.50	0.2910±0.0088	0.889/0.822
Faro	96	10–29	94	(97.9)	2	(2.1)	–	–	–		18.43±0.34	17.50	17	0.76±0.25	1.94±0.49	0.4509±0.0150	0.771/0.812
Guarda	80	10–36	78	(97.5)	1	(1.3)	1	(1.3)	–		18.41±0.42	17.00	17	1.92±0.27	6.03±0.53	0.2702±0.0129	0.775/0.813
Leiria	92	13–31	90	(97.8)	2	(2.2)	–	–	–		18.37±0.32	17.00	17	1.53±0.25	3.35±0.50	0.5552±0.0138	0.870/0.807
Lisboa	100	10–33	94	(94.0)	6	(6.0)	–	–	–		18.74±0.32	18.00	17	1.62±0.24	3.97±0.48	0.0662±0.0067	0.900/0.851
Madeira island	82	12–30	80	(97.6)	2	(2.4)	–	–	–		18.27±0.37	17.00	17	1.46±0.27	2.11±0.53	0.3155±0.0140	0.756/0.777
Portalegre	92	9–28	91	(98.9)	1	(1.1)	–	–	–		18.12±0.29	17.00	17	0.67±0.25	2.48±0.50	0.0871±0.0060	0.891/0.816
Porto	94	10–31	89	(94.7)	5	(5.3)	–	–	–		18.51±0.36	17.00	17	1.45±0.25	3.00±0.49	0.2446±0.0128	0.787/0.834
Santarém	90	13–26	90	(100.0)	–	–	–	–	–		17.98±0.26	17.00	17	1.33±0.25	2.00±0.50	0.5707±0.0114	0.778/0.795
Setúbal	80	11–32	75	(93.8)	5	(6.3)	–	–	–		19.27±0.45	18.00	17	1.25±0.27	2.09±0.53	0.6973±0.0128	0.875/0.882
Viana do Castelo	84	10–31	82	(97.6)	2	(2.4)	–	–	–		18.26±0.36	17.00	17	1.09±0.26	2.97±0.52	0.0242±0.0034*	0.738/0.858
Vila Real	92	10–30	87	(94.6)	5	(5.4)	–	–	–		18.52±0.36	17.00	17	1.45±0.25	2.76±0.50	0.6216±0.0140	0.804/0.778
Viseu	90	10–35	87	(96.7)	3	(3.3)	–	–	–		18.59±0.37	17.00	17	2.07±0.25	7.37±0.50	0.3411±0.0146	0.822/0.790
Portugal	1772	9–40	1717	(96.9)	53	(3.0)	1	(0.1)	1	(0.1)	18.44±0.08	17.61	17	1.41±0.06	4.02±0.12	0.3486±0.0146	0.810/0.815

* $P < 0.05$
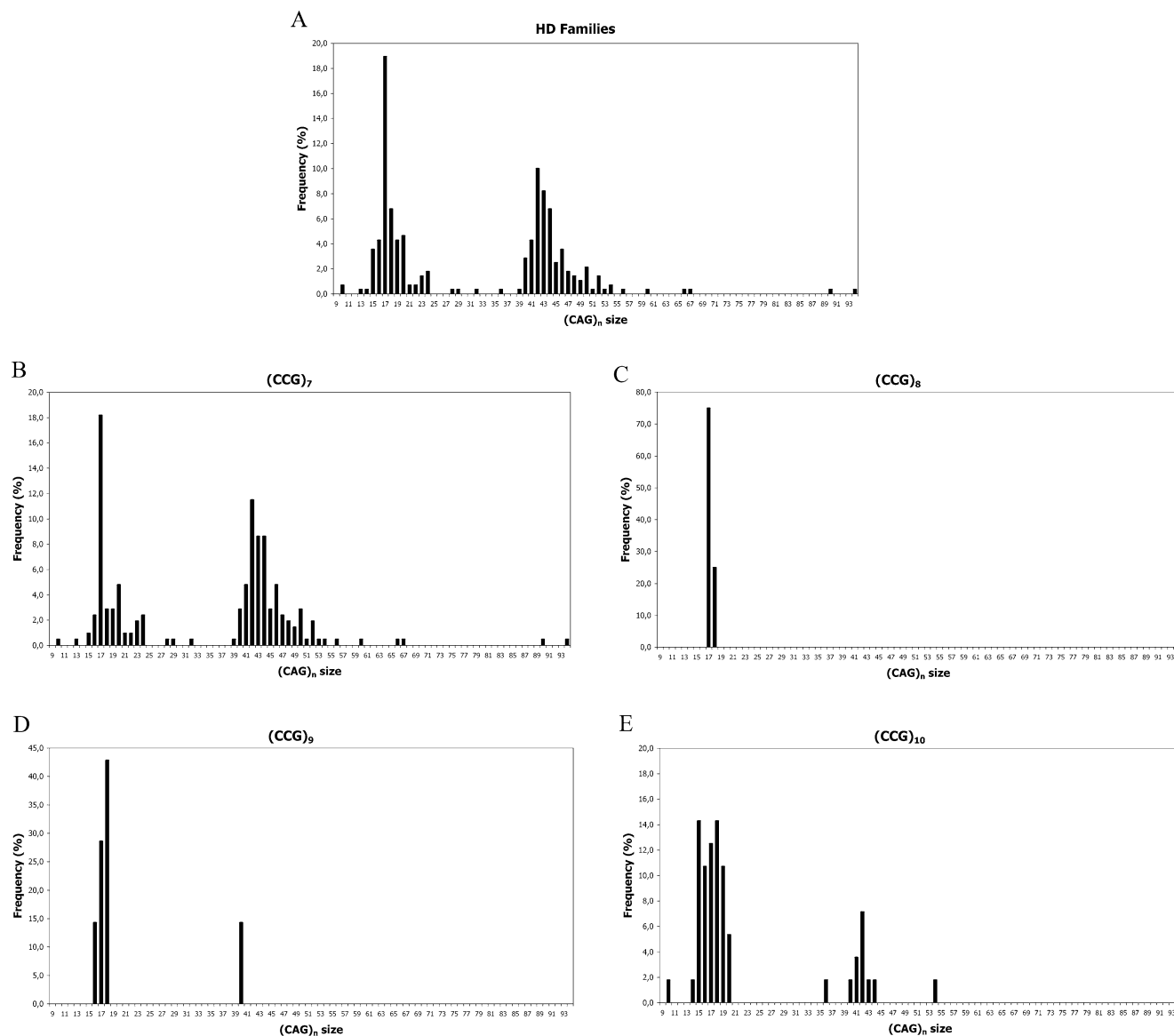
Fig. 2a–e Haplotype analysis in HD families. (**a**) Distribution of the CAG repeat size in the 140 HD families, each one represented by one randomly selected individual carrying the affected chromosome. Distribution of the CAG repeat length in families carrying the repeat 7-CCG (**b**), 8-CCG (**c**), 9-CCG (**d**) or 10-CCG (**e**)

Discussion

Maintenance of the frequency of mutant HD alleles (≥ 40 CAGs) in the population has been proposed to be due to a mutational bias towards the expansion of all sizes of normal alleles, but with a higher mutation rate for the longer normal alleles (Rubinsztein et al. 1994). The current analysis of the CAG repeat at the *HD* locus in the normal Portuguese population (1,772 chromosomes) is, to our knowledge, the largest control population study of this repeat and seems to show the previously proposed bias (Rubinsztein et al. 1994). The distribution of $(CAG)_n$ size in this population was similar to other described populations of Western European origin, showing a mode of 17 CAGs and a positive skewness (representing more alleles lying above the modal size) (Kremer et al. 1994; Rubinsztein et al. 1994; Squitieri et al. 1994). The Portuguese population was in Hardy–Weinberg equilibrium relative to the CAG repeat in the *HD* locus, showing that there was a random union of the gametes. Heterozygosity (0.810) was also similar to that of other European populations (Watkins et al. 1995; Andrés et al. 2002). The mean of the CAG repeat was also similar (Squitieri et al. 1994; Andrés et al. 2002). The $(CAG)_n$ ranged from 9 to 40 CAGs, and two expanded alleles were found. The presence of these two mutated alleles in a randomly selected sample of 1,000 individuals from our population could correspond to a higher prevalence of HD in the Portuguese population

Table 2 CCG repeat lengths in normal and expanded $(CAG)_n$ chromosomes in 140 Portuguese HD families

$(CCG)_n$ length in HD family	Normal $(CAG)_n$ chromosomes						Expanded $(CAG)_n$ chromosomes					
	Total		(< 27 CAGs)		(27–35 CAGs)		Total		(36–39 CAGs)		(> 39 CAGs)	
	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)
$(CCG)_7$	85	(61.2)	82	(60.3)	3	(100.0)	124	(87.9)	1	(50.0)	123	(88.5)
$(CCG)_8$	8	(5.8)	8	(5.9)	–	–	–	–	–	–	–	–
$(CCG)_9$	6	(4.3)	6	(4.4)	–	–	1	(0.7)	–	–	1	(0.7)
$(CCG)_{10}$	40	(28.8)	40	(29.4)	–	–	16	(11.3)	1	(50.0)	15	(10.8)

than described elsewhere. In a previous report, we had predicted it to be at least 2–5:10,0000 (Costa et al. 2003), i.e. similar to that of other European countries. The appearance of mutated alleles for HD in control populations has been reported only once, to the best of our knowledge, where a 39-CAG repeat was found among 600 control chromosomes (Kremer et al. 1994). Since it has been proposed that expanded HD chromosomes could result from a reservoir of normal unstable alleles with a particular haplotype, namely with the 7-CCG allele (Goldberg et al. 1993; Squitieri et al. 1994; Almqvist et al. 1995), the presence of two mutated alleles in our control population may potentially be explained by the relatively high frequency of class 2 alleles (3.0%) as compared to other reported studies (Goldberg et al. 1993, 1995; Kremer et al. 1994); however, the frequency of class 2 alleles observed in the regions where the mutated alleles were found (3.0% in the Azores and 1.3% in Guarda) was not higher than in the other regions, which may suggest that the origin of the mutated alleles is independent of the frequency of intermediate normal unstable alleles. Additionally, the total sample was shown to be homogeneous, without geographic clusterings, since the proportion of the four allele classes was similar in all regions, as was the distribution of CAG repeat size.

Thus, it is possible that the haplotypic environment of the CAG repeat in the *HD* gene is of great importance, both to its evolution and to the origin of newly expanded chromosomes. Haplotypes in 140 Portuguese families showed a distribution of CCG repeat size similar to that of other Western European populations (Squitieri et al. 1994). No significant effect of the size of the CCG repeat in the normal chromosome on age at onset was observed, which is in accordance with other studies (Vuillaume et al. 1998). Nevertheless, the carriers of a 7-CCG repeat associated with the expanded allele showed a lower age at onset when compared to those carrying a 10-CCG allele; however, given our sample size, the difference was not significant and was not associated with differences in mean $(CAG)_n$ size between the two groups. The normal alleles from the sample of independent Portuguese patients showed a distribution of the CAG repeat size similar to that observed in the control population. In our sample, as in others (Squitieri et al. 1994; Pramanik et al.

2000), the 8-CCG allele is a rare allele and was associated exclusively with normal chromosomes as well as showing the higher frequency of this allele (5.8%) reported until now. As in other populations studied (Andrew et al. 1994; Squitieri et al. 1994; Masuda et al. 1995; Pramanik et al. 2000): (1) the Portuguese families showed more than one HD founder haplotype (three different HD haplotypes were found to be associated with a 7, 9 or 10-CCG repeats); (2) the 7-CCG repeat was the most frequent allele, both in normal and in expanded alleles; (3) the 7-CCG repeat was preferentially associated with expanded chromosomes; (4) the mean CAG repeat size was higher in normal chromosomes containing 7 CCGs than in those with 10 CCGs; (5) indeed, all class 2 alleles had 7 CCGs, suggesting that these chromosomes may be more prone to full expansion; (6) the distribution of CAG repeat sizes among the 7-CCG haplotype was very similar to that of the total sample, and with the same bias towards larger CAG repeats, probably due to a more ancient age of this haplotype, as has been suggested (Rubinsztein et al. 1994). The same was not observed, however, for the CAG repeat distributions for any of the other CCG haplotypes, which showed very little variability. So, processes other than mutation bias and the reservoir of class 2 alleles could also underlie the evolution from normal size to expanded CAG repeats. As suggested for other triplet repeat disorders [such as myotonic dystrophy, fragile X syndrome (Paulson and Fischbeck 1996) and Friedreich ataxia (Cossée et al. 1997)], a stepwise model for evolution to full expansion may also occur in HD and could explain the origin of expanded chromosomes.

In conclusion, we suggest that (1) the HD mutation has multiple origins among the Portuguese population (three different haplotypes were associated with the disease); and (2) three mechanisms occurring at different times during evolution may have concurred, leading to evolution from normal CAGs to full expansion: first, a mutation bias towards larger alleles giving rise to a higher variability of normal chromosomes followed by a stepwise process that could explain the CAG distributions observed in the more recent haplotypes and, finally, a pool of intermediate (class 2) alleles (probably originated through the latter bias) more prone to give rise to reduced and fully penetrant HD alleles.

References

Almqvist EW et al (1995) Ancestral differences in the distribution of the Δ 2642 glutamic acid polymorphism is associated with varying CAG repeat lengths on normal chromosomes: insights into the genetic evolution of Huntington disease. Hum Mol Genet 4:207–214

Andrés AM, Lao O, Soldevila M, Calafell F, Bertranpetit J (2002) Dynamics of CAG repeat loci revealed by the analysis of their variability. Hum Mutat 21:61–70

Andrew SE et al (1993) The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. Nat Genet 4:398–403

Andrew SE, Goldberg YP, Theilmann J, Zeisler J, Hayden MR (1994) A CCG repeat polymorphism adjacent to the CAG repeat in the Huntington disease gene: implications for diagnostic accuracy and predictive testing. Hum Mol Genet 3:65–67

Cossée M et al (1997) Evolution of the Friedreich's ataxia trinucleotide repeat expansion: founder effect and premutations. Proc Natl Acad Sci USA 94:7452–7457

Costa MC et al (2003) Molecular diagnosis of Huntington disease in Portugal: implications for genetic counselling and clinical practice. Eur J Hum Genet 11:872–878

Duyao M et al (1993) Trinucleotide repeat length instability and age of onset in Huntington's disease. Nat Genet 4:387–392

Goldberg YP et al (1993) Molecular analysis of new mutations for Huntington's disease: intermediate alleles and sex of origin effects. Nat Genet 5:174–179

Goldberg YP et al (1995) Increased instability of intermediate alleles in families with sporadic Huntington disease compared to similar sized intermediate alleles in the general population. Hum Mol Genet 4:1911–1918

Guo SW, Thompson EA (1992) Performing the exact test of Hardy–Weinberg proportion for multiple alleles. Biometrics 48:361–372

Kremer B et al (1994) A worldwide study of Huntington's disease mutation: the sensitivity and specificity of measuring CAG repeats. N Engl J Med 330:1401–1406

Martin JB, Gusella JF (1986) Huntington's disease: pathogenesis and management. N Engl J Med 315:1267–1276

Masuda N, Goto J, Murayama N, Watanable M, Kondo R, Kanazawa I (1995) Analysis of triplet repeats in the Huntington gene in Japanese families affected with Huntington's disease. J Med Genet 32:701–705

Paulson HL, Fischbeck KH (1996) Trinucleotide repeats in neurogenetic disorders. Annu Rev Neurosci 19:79–107

Pramanik S et al (2000) Analysis of CAG and CCG repeats in Huntington gene among HD patients and normal populations of India. Eur J Hum Genet 8:678–682

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Hered 86:284–249

Rubinsztein DC 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

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning—a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Squitieri F et al (1994) DNA haplotype analysis of Huntington disease reveals clues to the origins and mechanisms of CAG expansion and reasons for geographic variations of prevalence. Hum Mol Genet 3:2103–2114

The American College of Medical Genetics, American Society of Human Genetics, Huntington Disease Genetic Testing Working Group (1998) Laboratory guidelines for Huntington disease genetic testing. Am J Hum Genet 62:1243–1247

The Huntington Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72:971–983

Vuillaume I, Vermersch P, Destée A, Petit H, Sablonnière B (1998) Genetic polymorphisms adjacent to the CAG repeat influence clinical at onset in Huntington's disease. J Neurol Neurosurg Psychiatry 64:758–762

Watkins WS, Bamshad M, Jorde LB (1995) Population genetics of trinucleotide repeat polymorphisms. Hum Mol Genet 4:1485–1491